

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 17, 2003, 19:37:48 ; Search time 2944 Seconds

(without alignments)
10676.301 Million cell updates/sec

Title: US-09-972-758a-1

Perfect score: 1080
Sequence: 1 atgcgcgacgcatctctgtc.....tttccaagtgttgagactag 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GeneBml:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1076.8	99.7	3624	9	AB021179	AB021179 Homo sapi
2	1076.8	99.7	100836	2	AC087298	AC087298 Homo sapi
3	1075.2	99.6	2178	9	AK023624	AK023624 Homo sapi
4	1073.6	99.4	2086	9	BC006460	BC006460 Homo sapi
5	1052.8	97.5	142326	2	AC024047	AC024047 Homo sapi
6	840.8	77.9	3402	10	AY090614	AY090614 Homo sapi
7	840.8	77.9	111002	10	AL731805	AL731805 Mouse DNA
8	833.8	77.2	1488	10	BC022111	BC022111 Mus muscu
9	793.8	73.5	175120	2	AC120950	AC120950 Rattus no
10	355.2	32.9	461	6	AX31371	AX31371 Sequence
11	174.8	16.2	1238	9	BC025970	BC025970 Homo sapi
12	174.8	16.2	1330	9	AK056946	AK056946 Homo sapi
13	174.8	16.2	116035	2	AC008105	AC008105 Homo sapi
14	173.2	16.0	22048	9	AK096785	AK096785 Homo sapi
15	172	15.9	223442	10	AL662804	AL662804 Mouse DNA
16	103	9.5	167636	2	AC107097	AC107097 Rattus no
17	99.4	9.2	82400	10	AC090495	AC090495 Genomic s
18	99.2	9.2	195673	2	AL807807	AL807807 Mus muscu
19	87.4	8.1	198228	2	AC102097	AC102097 Mus muscu
20	87	8.1	207683	2	AC098712	AC098712 Mus muscu
21	86.2	8.0	283816	2	AC124587	AC124587 Mus muscu
22	85.8	7.9	194334	2	AC101527	AC101527 Mus muscu
23	84.8	7.9	248037	2	AC121887	AC121887 Mus muscu
24	84.4	7.8	195690	10	AC098726	AC098726 Mus muscu
25	84.2	7.8	155662	2	AC112388	AC112388 Rattus no
26	84	7.8	180418	2	AC106674	AC106674 Rattus no
27	83.4	7.7	231853	2	AC093352	AC093352 Mus muscu
28	83	7.7	225045	2	AL833803	AL833803 Mus muscu
29	82.8	7.7	318930	2	AC073495	AC073495 Mus muscu
30	80.4	7.4	168210	2	AC116700	AC116700 Mus muscu
31	79.6	7.4	184865	2	AL807824	AL807824 Mus muscu
32	79.2	7.3	272545	2	AC090533	AC090533 Mus muscu
33	78.6	7.3	165197	2	AC112538	AC112538 Rattus no
34	78.2	7.2	203982	9	AC067854	AC067854 Homo sapi
35	77.8	7.2	234817	10	AL663048	AL663048 Mouse DNA
36	77.6	7.2	185460	2	AC115552	AC115552 Rattus no
37	77.4	7.2	180850	2	AC107828	AC107828 Mus muscu
38	75.4	7.0	287323	2	AC093363	AC093363 Mus muscu
39	75.2	7.0	97348	10	AF091216	AF091216 Mus muscu
40	74.4	6.9	256346	2	AC105970	AC105970 Mus muscu
41	73.8	6.8	126357	2	AC121480	AC121480 Rattus no
42	73.8	6.8	201271	2	AC124194	AC124194 Mus muscu
43	73.8	6.8	206497	10	AC084389	AC084389 Mus muscu
44	73.2	6.8	186400	2	AC095843	AC095843 Rattus no
45	73.2	6.8	227386	2	AC125235	AC125235 Mus muscu

ALIGNMENTS

RESULT 1
AB021179 3624 bp mRNA linear PRI 05-DEC-2000
LOCUS Homo sapiens mRNA for HEXIM1 protein, complete cds.
DEFINITION
AB021179
ACCESSION
AB021179.1 GI:4062855
VERSION
HEXIM1; HEXIM1 protein.
KEYWORDS
Homo sapiens coronary artery smooth muscle cell cDNA to mRNA.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
Kusuhara,M., Nagasaki,K., Kimura,K., Maass,N., Manabe,T.,
Ishikawa,S., Aikawa,M., Miyazaki,K. and Yamaguchi,K.
TITLE
Cloning of hexamethylene-bis-acetamide-inducible transcript,

Pred. No. is the number of results predicted by chance to have a


```

Db 70725 GCTGCTGCTGTCAGGAAGAGCTGAACCTTGAGCCGCCCCAGCGCGAGAGCGGGTG 70666
QY 121 CCGGAGGAGAGAGTAGAGTGGCAATCGAGAGCGTCCCAATTGGTGGCCGCTCCGGG 180
Db 70665 CCGGAGGAGAGAGTAGAGTGGCAATCGAGAGCGTCCCAATTGGTGGCCGCTCCGGG 70606
QY 181 CCGGAGGAGGAGAGAGCTGGATCCCAACCACTCCCTTGACAGACCAGCCAGCTGTCCA 240
Db 70605 CCGGAGGAGGAGAGAGAGCTGGATCCCAACCACTCCCTTGACAGACCAGCCAGCTGTCCA 70546
QY 241 GAATCTAGCTGCTCTAGAGAGAGGCGGAGAAAGGCGCAAGTGGGAGAGACTGTCCTGGC 300
Db 70545 GAATCTAGCTGCTCTAGAGAGAGGCGGAGAAAGGCGCAAGTGGGAGAGACTGTCCTGGC 70486
QY 301 GCGGACTTCCCGCCCGCGAGAGTGGAAACCGAGCCCGAGCGGAGTGTCTGCGCCAG 360
Db 70485 GCGGACTTCCCGCCCGCGAGAGTGGAAACCGAGCCCGAGCGGAGTGTCTGCGCCAG 70426
QY 361 CTTGTCTATGACTCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAGAG 420
Db 70425 CTTGTCTATGACTCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAGAG 70366
QY 421 TGGGAGACAGACAGACAGACAGCTGGGGGAGAAAGAAACATAGAGAGCCCTCCCAAGAG 480
Db 70365 TGGGAGACAGACAGACAGACAGCTGGGGGAGAAAGAAACATAGAGAGCCCTCCCAAGAG 70306
QY 481 AAGCGCATTTGGAAACCGTACTACAAAGTGAAGTGGGAGAGAGAAAGAAAGTTCCAGCAG 540
Db 70305 AAGCGCATTTGGAAACCGTACTACAAAGTGAAGTGGGAGAGAGAGAAAGAAAGTTCCAGCAG 70246
QY 541 AAACAGAGGCTTTCAGAGCTTCAAGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGT 600
Db 70245 AAACAGAGGCTTTCAGAGCTTCAAGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGT 70186
QY 601 GCGCCCTATACACACAGAGCTTCCATGATGATCAAGACAGAGAGAGCCGAGATC 660
Db 70185 GCGCCCTATACACACAGAGCTTCCATGATGATCAAGACAGAGAGAGCCGAGATC 70126
QY 661 AAACCGGCGCTTCTACTCCAAAGCGGGCGCCGCAATCCGAGACAGACAGCGATGACGAC 720
Db 70125 AAACCGGCGCTTCTACTCCAAAGCGGGCGCCGCAATCCGAGACAGACAGCGATGACGAC 70066
QY 721 TTCATGAGAAAGAGGGGTAGAGAGATGGGGGACAGATGGATGGGAGGGAGCGGCAC 780
Db 70065 TTCATGAGAAAGAGGGGTAGAGAGATGGGGGACAGATGGGATGGGAGGGAGCGGCAC 70006
QY 781 GAGTTTCTGACGCGGACTTTCGAGAGCTACGAGCGGTACACACAGAGAGACCTGCAG 840
Db 70005 GAGTTTCTGACGCGGACTTTCGAGAGCTACGAGCGGTACACACAGAGAGACCTGCAG 69946
QY 841 AACATGAGCAGCAGAGCTCATCAAGAGTACTGGAACCTGGAGAACTGCTTCGCGC 900
Db 69945 AACATGAGCAGCAGAGCTCATCAAGAGTACTGGAACCTGGAGAACTGCTTCGCGC 69886
QY 901 AAGGAGAGAGAGAAACACCGGCTGGGAGAGAGAAACGCGTGGTGGCCAGACGCG 960
Db 69885 AAGGAGAGAGAGAAACACCGGCTGGGAGAGAGAAACGCGTGGTGGCCAGACGCG 69826
QY 961 CGTGTGCGGAGCTGAGAGCTGAGAGTGAACCGGCTGCGCGAGAAACCTCCAGCTGTG 1020
Db 69825 CGTGTGCGGAGCTGAGAGCTGAGAGTGAACCGGCTGCGCGAGAAACCTCCAGCTGTG 69766
QY 1021 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 69765 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69706

```

RESULT 3
 AK023624 2178 bp mRNA linear PRI 01-AUG-2002
 LOCUS
 DEFINITION Homo sapiens CDNA FLJ13562 fis, clone PLACE1008080, highly similar
 to Homo sapiens mRNA for HEXIM1 protein.
 ACCESSION AK023624

```

VERSION AK023624.1 GI:10435606
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens Placenta CDNA to mRNA, clone_1lib:PLACE1
        clone:PLACE1008080.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
    Nishikawa,T., Naei,K., Sugano,S., Shiratori,A., Sudo,H.,
    Magatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
    Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
    Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Murakami,K.,
    Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
    Nakamura,Y., Nagahara,K., Masuhara,Y., Ninomiya,K. and Iwayanagi,T.
    NEDO human cDNA sequencing project
  2 (bases 1 to 2178)
  Unpublished
REFERENCE
  1 Isogai,T. and Otsuki,T.
    Direct Submission
    Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
    Genomics Laboratory: 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
    (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT
  NEDO human cDNA sequencing project supported by Ministry of
  International Trade and Industry for Biotechnology; cDNA library
  sequencing; Research Association for Biotechnology; cDNA library
  construction, 5'- & 3'-end one pass sequencing and clone selection;
  Helix Research Institute (supported by Japan Key Technology Center
  etc.) and Department of Virology, Institute of Medical Science,
  University of Tokyo.
FEATURES
  source
    1..2178
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="PLACE1008080"
    /tissue_type="Placenta"
    /clone_1lib="PLACE1"
    /note="Cloning vector: pME18SFL3"
BASE COUNT
  525 a 513 c 678 g 462 t
ORIGIN
  Query Match 99.6%; Score 1075.2; DB 9; Length 2178;
  Best Local Similarity 99.7%; Pred. No. 6.5e-169;
  Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
  1 ATGCCGAGGCCATTCTTGTTCAGAAATATCAACACGACCTCAAACTAGCACTGTACAGT 60
  716 ATGGCCGAGGCCATTCTTGTTCAGAAATATCAACACGACCTCAAACTAGCACTGTACAGT 775
  61 GCTGCTGCTGTCACAGAGAGCTGAACCTTGAGCCGCCCCAGCGCGGAGAGAGGGGTG 120
  776 GCTGCTGCTGTCACAGAGAGCTGAACCTTGAGCCGCCCCAGCGCGGAGAGAGGGGTG 835
  121 CCCGAGAGAGCAGTAGAGTGGCAATCGAGAGAGCTTCCCAATGGTGGTCCGCGGG 180
  836 CCCGAGAGAGCAGTAGAGTGGCAATCGAGAGAGCTTCCCAATGGTGGTCCGCGGG 895
  181 CCGAGAGGGGAGAGAGAGCTGGAATCCCAACCACTCTCCCTGACAGACCGAGCTGTGCA 240
  896 CCGAGAGGGGAGAGAGAGCTGGAATCCCAACCACTCTCCCTGACAGACCGAGCTGTGCA 955
  241 GAATCTAGCTGCTCTAGAGAGAGGCGGAGAAAGGCGCAAGTGGGAGAGAGCTGTCGCTGGC 300
  956 GAATCTAGCTGCTCTAGAGAGAGGCGGAGAAAGGCGCAAGTGGGAGAGAGCTGTCGCTGGC 1015
  1016 GCGGACTTCCCGCCCGCGAGAGAGTGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075
  361 CTTGTCTATGACTCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAGAG 420
  1076 CTTGTCTATGACTCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAGAG 1135

```

```

OY 421 TGGGACAGACAGACAGACAGCTGGGGAAGAAAAACATTAAGACGCCCGCTCCAGAAG 480
    |||||||
Db 1136 TGGGACAGACAGACAGACAGCTGGGGAAGAAAAACATTAAGACGCCCGCTCCAGAAG 1195
OY 481 AAGCGCATTTGGAACCGCTACTACAAAGTGAACCTGGGAAGAAAAAGTTTCAGCAG 540
    |||||||
Db 1196 AAGCGCATTTGGAACCGCTACTACAAAGTGAACCTGGGAAGAAAAAGTTTCAGCAG 1255
OY 541 AACAGACGCTTCGAGCTTCAAGGATCCGACCCGAGATGTTGCCCAAGGCCAGCGGTC 600
    |||||||
Db 1256 AACAGACGCTTCGAGCTTCAAGGATCCGACCCGAGATGTTGCCCAAGGCCAGCGGTC 1315
OY 601 GCGCCCTATTAACACACAGCAGTTCCTCATGTGATGATCAGACACAGAGAGAGCGGATCTC 660
    |||||||
Db 1316 GCGCCCTATTAACACACAGCAGTTCCTCATGTGATGATCAGACACAGAGAGAGCGGATCTC 1375
OY 661 AAAACGGGCTGTACTCCAGCGGGCGCGCCCAATCCGAGACACAGCAGCGATGACAC 720
    |||||||
Db 1376 AAAACGGGCTGTACTCCAGCGGGCGCGCCCAATCCGAGACACAGCAGCGATGACAC 1435
OY 721 TTTCATGGAAGAAGGGGGTGAGAGAGATGGGGGACACCATGGGATGGAGGGGACGCGAC 780
    |||||||
Db 1436 TTTCATGGAAGAAGGGGGTGAGAGAGATGGGGGACACCATGGGATGGAGGGGACGCGAC 1495
OY 781 GAGTTTTCGACGCGGAGCTTCCTCGAGAGAGTACGAGCGGTACACACAGAGAGAGCGTGCAG 840
    |||||||
Db 1496 GAGTTTTCGACGCGGAGCTTCCTCGAGAGAGTACGAGCGGTACACACAGAGAGAGCGTGCAG 1555
OY 841 AACATGACCAAGCAGACGAGCTCATCAAGAGAGTACCTGGAATCGAGAACTGCTCTCGCGC 900
    |||||||
Db 1556 AACATGACCAAGCAGACGAGCTCATCAAGAGAGTACCTGGAATCGAGAACTGCTCTCGCGC 1615
OY 901 ATGGAGAGCAAGCAACACGGCTGGCGCTGGAGACCAAGCGCTGGTGGGACAGACGCG 960
    |||||||
Db 1616 ATGGAGAGCAAGCAACACGGCTGGCGCTGGAGACCAAGCGCTGGTGGGACAGACGCG 1675
OY 961 CGTGTGCGGAGCTGGAGCTGGAGCTGGAGCGGCTGGCGCGGAGAACCTCCAGCTGCTG 1020
    |||||||
Db 1676 CGTGTGCGGAGCTGGAGCTGGAGCTGGAGCGGCTGGCGCGGAGAACCTCCAGCTGCTG 1735
OY 1021 ACCGAGAACGAACTGACACCGGACAGAGAGCGCGCGCTTCCCAAGTTTGGAGACTAG 1080
    |||||||
Db 1736 ACCGAGAACGAACTGACACCGGACAGAGAGCGCGCGCTTCCCAAGTTTGGAGACTAG 1795

```

```

RESULT 4
BC006460 2086 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, HMBA-Inducible, clone MGC:1880 IMAGE:353529, mRNA,
ACCESSION BC006460
VERSION BC006460.1 GI:13623668
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2086)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Rubin Laboratory
cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca

```

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Neta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carlie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Iliisa Prabh, Parvaneh Seedi, Jacqueline Schein, Duane Smalls, Michael Stahl, Loraine Spence, Jeff Stolt, Michael Thorne, Miranada Tsai, Natalja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL, Plate: 8 Row: d Column: 1
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4062855.

FEATURES

source

```

1..2086
/organism="Homo sapiens"
/db_xref="locusid:10614"
/db_xref="taxon:9606"
/clone="MGC:1880 IMAGE:353529"
/tissue_type="Lung, small cell carcinoma"
/clone_id="NIH-MGC_7"
/lab_host="DH10B-R"
/note="Vector: pCIB7"
606..1685
/codon_start=1
/product="HMBA-inducible"
/protein_id="AAH06460.1"
/db_xref="gi:13623668"

```

CDS

```

BASE COUNT 510 a 505 c 618 g 453 t
ORIGIN
Query Match 99.4%; Score 1073.6; DB 9; Length 2086;
Best Local Similarity 99.6%; Pred. No. 1.2e-168;
Matches 1076; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 ATGGCGAGACCATTTGTGCAAGATATACACCGAGCTCAACTAGACAGTGTACAGT 60
    |||||||
Db 606 ATGGCGAGACCATTTGTGCAAGATATATACACCGAGCTCAACTAGACAGTGTACAGT 665
OY 61 GCTGCTGCTGCCAGAGAGAGCTGAACCTGAGCGCCCGCCAGGCGGAGAGCGGCTG 120
    |||||||
Db 666 GCTGCTGCTGCCAGAGAGAGCTGAACCTGAGCGCCCGCCAGGCGGAGAGCGGCTG 725
OY 121 CCCGAGAGAGACAGTAGTGCATCGAATCGAGAGCGTTCCCGCAAGTTGGTGGCCGCGGG 180
    |||||||
Db 666 GCTGCTGCTGCCAGAGAGAGCTGAACCTGAGCGCCCGCCAGGCGGAGAGCGGCTG 725
OY 726 CCCGAGAGAGACAGTAGTGCATCGAATCGAGAGCGTTCCCGCAAGTTGGTGGCCGCGGG 785
    |||||||
Db 726 CCCGAGAGAGACAGTAGTGCATCGAATCGAGAGCGTTCCCGCAAGTTGGTGGCCGCGGG 785
OY 181 CCGAGAGGAGAGAGAGAGCTGGAATCCCAACCACTGCTTGGAGAGACCGAGCGGTGCA 240
    |||||||
Db 786 CCGAGAGGAGAGAGAGAGCTGGAATCCCAACCACTGCTTGGAGAGACCGAGCGGTGCA 845
OY 241 GAATCTACCTGCTGAGAGAGAGCGGAGAGGCGCGAGATGGGAGAGTCTGCGCTGGC 300
    |||||||
Db 846 GAATCTACCTGCTGAGAGAGAGCGGAGAGGCGCGAGATGGGAGAGTCTGCGCTGGC 905
OY 301 GCGCAGCTCCGCGCGCGCGGAGAGTGAACCGCGCGCGGAGCGGAGTGTGCGCCAG 360
    |||||||
Db 906 GCGCAGCTCCGCGCGCGCGGAGAGTGAACCGCGCGCGGAGCGGAGTGTGCGCCAG 965
OY 361 CTTTGTATGATCTCGAGAGCGGAGTGAAGTTGGGGCTCTGCGCGAGGGGGGGAAGAGAG 420
    |||||||
Db 966 CTTTGTATGATCTCGAGAGCGGAGTGAAGTTGGGGCTCTGCGCGAGGGGGGGAAGAGAG 1025
OY 421 TGGGACAGACAGACAGACAGCTGGGGAAGAAAAACATTAAGACGCCCGCTCCAGAAG 480
    |||||||

```

```

Db      1026 TGGGAGCAGCAGACAGCTGGGGAGAGAAAAAATAGAGAGCCCTGCAAGAG 1085
QY      481 AAGCGCATTTGAAACCGTACTACAGCTGAACTGGGAGAGAGAAAAAGTTGACGAG 540
Db      1086 AAGCGCATTTGAAACCGTACTACAGCTGAACTGGGAGAGAGAAAAAGTTGACGAG 1145
QY      541 AAACAGAGCCTTGCAGCTTCAGAGATCCAGCCGAGATGTTGCCAAGAGGCCAGCCGTC 600
Db      1146 AAACAGAGCCTTGCAGCTTCAGAGATCCAGCCGAGATGTTGCCAAGAGGCCAGCCGTC 1205
QY      601 GCGCCCTATACACCGAGCTTCCTCATGATGATCAGACAGAGAGAGCCGAGATC 660
Db      1206 GCGCCCTATACACCGAGCTTCCTCATGATGATCAGACAGAGAGAGCCGAGATC 1265
QY      661 AAACCGGCGCTTACTTCAAGCGCGCGCCGCAATCCGACACACAGCAGTGAAGAG 720
Db      1266 AAACCGGCGCTTACTTCAAGCGCGCGCCGCAATCCGACACACAGCAGTGAAGAG 1325
QY      721 TTATGAGAGAGAGGGGTGAGAGAGATGGGGGCGACGATGGATGGAGGGGCGACGAGC 780
Db      1326 TTATGAGAGAGAGGGGTGAGAGAGATGGGGGCGACGATGGATGGAGGGGCGACGAGC 1385
QY      781 GACTTCTGCGAGGGGACTCTGCGAGAGCTAGAGGGGTACCAACAGAGAGAGCCTGACG 840
Db      1386 GACTTCTGCGAGGGGACTCTCTCGAGAGCTAGAGGGGTACCAACAGAGAGCCTGACG 1445
QY      841 AACATGAGCAGAGAGAGCTCATCAAGAGATGCTGAACTGGAGAGAGTCCCTCTGCGCG 900
Db      1446 AACATGAGCAGAGAGAGCTCATCAAGAGATGCTGAACTGGAGAGAGTCCCTCTGCGCG 1505
QY      901 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Db      1506 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1565
QY      961 CGTGTGCGGAGAGTGGAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db      1566 CGTGTGCGGAGAGTGGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1625
QY      1021 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db      1626 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1685

RESULT 5
AC024047/c 142326 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 17 clone RP11-403G3, WORKING DRAFT
DEFINITION
AC024047
AC024047.3 GI:8570008
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 142326)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 142326)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jun 16, 2000 this sequence version replaced gi:7109599.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0403G03

```

```

----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid, 0%
Chemistry: Dye-terminator Big Dye, 100% of reads
Chemistry: Dye-terminator Big Dye, 0% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 127942 bases at least Q40
Consensus quality: 13308 bases at least Q30
Consensus quality: 135734 bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 139626; sum-of-ctnigs
Quality coverage: 4.06 in Q20 bases; agarose-fp
Quality coverage: 4.33 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1575: contig of 1575 bp in length
1576 1675: gap of unknown length
1676 3114: contig of 1438 bp in length
3115 3214: gap of unknown length
3215 5389: contig of 2175 bp in length
5390 5489: gap of unknown length
5490 7129: contig of 1640 bp in length
7130 7229: gap of unknown length
7230 9240: contig of 2011 bp in length
9241 9340: gap of unknown length
9341 11177: contig of 1837 bp in length
11178 11277: gap of unknown length
11278 13113: contig of 1836 bp in length
13114 13213: gap of unknown length
13214 16526: contig of 3313 bp in length
16527 16626: gap of unknown length
16627 19864: contig of 3238 bp in length
19865 19964: gap of unknown length
19965 22667: contig of 2703 bp in length
22668 22767: gap of unknown length
22768 26650: contig of 3883 bp in length
26651 26750: gap of unknown length
26751 29303: contig of 2553 bp in length
29304 32073: contig of 2670 bp in length
32074 32173: gap of unknown length
32174 37141: contig of 4968 bp in length
37142 37241: gap of unknown length
37242 40838: contig of 3497 bp in length
40839 45166: gap of unknown length
45167 45266: contig of 4328 bp in length
45267 49501: gap of unknown length
49502 49601: contig of 4235 bp in length
49602 55647: gap of unknown length
55648 55747: contig of 6046 bp in length
55748 61532: contig of 5785 bp in length
61533 61632: gap of unknown length
61633 68105: contig of 6373 bp in length
68106 68106: gap of unknown length
68107 75211: contig of 7106 bp in length
75212 81560: contig of 6249 bp in length
81561 81660: gap of unknown length
81661 88524: contig of 6864 bp in length
88525 88624: gap of unknown length
88625 95519: contig of 6895 bp in length
95520 105148: gap of unknown length
105149 105248: contig of 9529 bp in length
105249 115269: contig of 10021 bp in length

```



```

FEATURES
  source
    1. 142326
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="17"
    /clone="RP11-403g3"
BASE COUNT      32503 a 37098 c 37256 g 32736 t 2733 others
ORIGIN
Query Match      97.5%: Score 1052.8; DB 2; Length 142326;
Best Local Similarity 99.6%: Pred. No. 2e-165;
Matches 1076; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 ATGGCCGAGCCATCTTGTCAAGATATCAACACAGCCTCAAACTAGCACTGTACAGT 60
    |||||||
Db 108505 ATGGCCGAGCCATCTTGTCAAGATATCAACACAGCCTCAAACTAGCACTGTACAGT 108446
    |||||||

QY 61 GCTCTGCTGTCCAGGAAGAGCTTAACCTGAGCGCCGCCAGCGCGGAGAGCGGCTG 120
    |||||||
Db 108445 GCTCTGCTGTCCAGGAAGAGCTTAACCTGAGCGCCGCCAGCGCGGAGAGCGGCTG 108386
    |||||||

QY 121 CCCGAGGAGCACTAGGTGGCAATCGAGAGCTTCCGCCGAGTGGGGGCGCTCCGCGG 180
    |||||||
Db 108385 CCCGAGGAGCACTAGGTGGCAATCGAGAGCTTCCGCCGAGTGGGGGCGCTCCGCGG 108326
    |||||||

QY 181 CCGGAGGAGGGAAGGAGCCTTGAAATCCCAACACCTCCCTTGCAGAGCCAGCCTGTCCA 240
    |||||||
Db 108325 CCGGAGGAGGGAAGGAGCCTTGAAATCCCAACACCTCCCTTGCAGAGCCAGCCTGTCCA 108266
    |||||||

QY 241 GAATCTAGCTCCCTGAGAGAGAGGCGGAGAAAGGCCAAGATGGGGGACGACTCTCCGCTGCG 300
    |||||||
Db 108265 GAATCTAGCTCCCTGAGAGAGAGGCGGAGAAAGGCCAAGATGGGGGACGACTCTCCGCTGCG 108206
    |||||||

QY 301 GGCAGCTTCCCGCGCGCGGAGAGTGAACCGAGCGCGCGGAGCTCTCGGCCAG 360
    |||||||
Db 108205 -GGCAGCTTCCCG -CGCGCGGAGAGAGTGAACCGAGCGCGCGGAGCTCTCGGCCAG 108148
    |||||||

QY 361 CCTTGTATGACTCTCGAGGCGAGTGAAGTTGGGGGCTCTCGCGCGGAGGAGAG 420
    |||||||
Db 108147 CCTTGTATGACTCTCGAGGCGAGTGAAGTTGGGGGCTCTCGCGCGGAGGAGAGAG 108088
    |||||||

QY 421 TGGGGAAGAGGAGAGAGAGAGTGGGGAAGAAAAAATAAGAGAGCGCCGCTCCAAGAG 480
    |||||||
Db 108087 TGGGGAAGAGGAGAGAGAGAGTGGGGAAGAAAAAATAAGAGAGCGCCGCTCCAAGAG 108028
    |||||||

QY 481 AAGGCGATTGGAAACGCTACTACAGTGAAGTGGGGAAGGAAAGAAAAAGTTGAGAGAG 540
    |||||||
Db 108027 AAGGCGATTGGAAACGCTACTACAGTGAAGTGGGGAAGGAAAGAAAAAGTTGAGAGAG 107968
    |||||||

QY 541 AAACAGAGCCTTCGAGGCTTCAAGGATCCAGCCGAGATGTTGGCCAAAGGGCCAGCCGATC 600
    |||||||
Db 107967 AAACAGAGCCTTCGAGGCTTCAAGGATCCAGCCGAGATGTTGGCCAAAGGGCCAGCCGATC 107908
    |||||||

QY 601 GCGGCCATTAACACACAGAGAGTTCCTCATGATATATACAGACCGAGAGAGCCGATCTC 660
    |||||||
Db 107907 GCGGCCATTAACACACAGAGAGTTCCTCATGATATATACAGACCGAGAGAGCCGATCTC 107848
    |||||||

QY 661 AAAACCGGCTGTACTCTCAAGCGGGCGCGCCAAATCCGAGCAGACAGAGGATGAGAGAG 720
    |||||||
Db 107847 AAAACCGGCTGTACTCTCAAGCGGGCGCGCCAAATCCGAGCAGACAGAGGATGAGAGAG 107788
    |||||||

QY 721 TTCAATGGAAGAGGGGTGAAGAGATGGGGGAGCGATGGAGTGGAGGGGAGCGGAGC 780
    |||||||
Db 107787 TTCAATGGAAGAGGGGTGAAGAGATGGGGGAGCGATGGAGTGGAGGGGAGCGGAGC 107728
    |||||||

QY 781 GAGTTTGTGAGCGCGGACTTCTGCGAGACGTACGAGGGGTACACACAGGAGAGCCTTCAG 840
    |||||||
Db 107727 GAGTTTGTGAGCGCGGACTTCTGCGAGACGTACGAGGGGTACACACAGGAGAGCCTTCAG 107668
    |||||||

```

```

QY 841 AACATGACGACGAGGAGCTCATCAAGAGTACCTGGAGACTGGAGAAAGTGCCTCCGCC 900
    |||||||
Db 107667 AACATGACGACGAGGAGCTCATCAAGAGTACCTGGAGAAAGTGCCTCCGCC 107608
    |||||||

QY 901 ATGGAGGACGAGAACCAACCGGCTCGGCTGGAGAGCAACCGCGTGGGTGGAGAGCAGCGC 960
    |||||||
Db 107607 ATGGAGGACGAGAACCAACCGGCTCGGCTGGAGAGCAACCGCGTGGGTGGAGAGCAGCGC 107548
    |||||||

QY 961 CGTGTGCGGAGGCTGGAGAGCTGGAGACCGGCTCGCGCGCGGAGAACCTTCAGCTGCTG 1020
    |||||||
Db 107547 CGTGTGCGGAGGCTGGAGAGCTGGAGACCGGCTCGCGCGCGGAGAACCTTCAGCTGCTG 107488
    |||||||

QY 1021 ACCGAGAACGAACTGCACCCGCGACAGAGAGCGACCGCCGCTTCCCACTTGGAGACTAG 1080
    |||||||
Db 107487 ACCGAGAACGAACTGCACCCGCGACAGAGAGCGACCGCCGCTTCCCACTTGGAGACTAG 107428
    |||||||

```

```

RESULT 6
LOCUS      AY090614
DEFINITION Mus musculus cardiac lineage protein 1 (Clp1) mRNA, complete cds.
ACCESSION AY090614
VERSION    AY090614.1 GI:20135640
KEYWORDS
SOURCE
  ORGANISM
    Mus musculus.
    Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 3402)
  Huang, F., Wagner, M. and Siddiqui, M. A.O.
  Structure, expression, and functional characterization of the mouse
  Clp-1 gene
  JOURNAL   Gene 292 (1-2), 245-259 (2002)
  MEDLINE   22114986
  PUBMED    12119119
REFERENCE
  2 (bases 1 to 3402)
  Huang, F., Wagner, M. and Siddiqui, M.
  Direct Submission
  JOURNAL   Submitted (19-MAR-2002) Anatomy and Cell Biology, State University
  of New York Health Science Center, 450 Clarkson Avenue, Brooklyn,
  NY 11203, USA
FEATURES
  source
    1..3402
    /organism="Mus musculus"
    /strain="129/SvJ"
    /db_xref="taxon:10090"
    1..3402
    /gene="Clp1"
    596..1666
    /gene="Clp1"
    /note="nuclear protein; CLP-1"
    /codon_start=1
    /product="cardiac lineage protein 1"
    /protein_id="AA09026.1"
    /translation="MAEPLILTEHOHOPSTNCTGAAYVHEHTSERPSAEPPKED
    SRMSRSALDSGSRPOEGEGKHLQPLPIQTNACPLSLBKEGKONKEDISTGSA
    SPSEAGEPMSSESLVOPGHDEATKQEAIPAAGGEPPWQOQRLKKHRRPSSKKRH
    WKPYKLTWEKKKFKDEKQSLIRASRYAEKFAAGQVAPYNTQFLMDHDOEPDLK
    TGIYPKRAAAKSDTSDDEVEEAGEEDGSDGGGSEFLQRFSEYERYHAESTL
    QNMSKELIKEYLELEKLSRKEDENNRLRLIESKRLGGVDARVRELELEIDRURENL
    QLTLENLHQOERAPLSKFGD"
BASE COUNT      920 a 771 c 890 g 821 t
ORIGIN
Query Match      77.9%: Score 840.8; DB 10; Length 3402;
Best Local Similarity 86.9%: Pred. No. 4.7e-130;
Matches 939; Conservative 0; Mismatches 132; Indels 9; Gaps 1;

```


- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

1	1115	1114	contig of 1114 bp in length
*	1115	1215	2688: contig of 1474 bp in length
*	2689	2788	gap of unknown length
*	2789	4582	contig of 1794 bp in length
*	4583	4682	gap of unknown length
*	4683	6524	contig of 1842 bp in length
*	6525	6625	gap of unknown length
*	6625	8343	contig of 1718 bp in length
*	8343	8442	gap of unknown length
*	8443	11551	contig of 3109 bp in length
*	11552	11651	gap of unknown length
*	13500	13499	contig of 1848 bp in length
*	13500	13599	gap of unknown length
*	13600	14781	contig of 1182 bp in length
*	14782	14681	gap of unknown length
*	14882	17066	contig of 2185 bp in length
*	17067	17166	gap of unknown length
*	17167	19189	contig of 2023 bp in length
*	19189	19290	gap of unknown length
*	19290	20786	contig of 1497 bp in length
*	20787	20886	gap of unknown length
*	20887	22783	contig of 1857 bp in length
*	22784	22883	gap of unknown length
*	22884	24379	contig of 1496 bp in length
*	24380	24479	gap of unknown length
*	24480	25004	contig of 1425 bp in length
*	25005	26004	gap of unknown length
*	26005	26872	contig of 2568 bp in length
*	26573	26872	gap of unknown length
*	26873	31006	contig of 2334 bp in length
*	31007	31106	gap of unknown length
*	31107	33737	contig of 2621 bp in length
*	33738	33837	gap of unknown length
*	33838	35985	contig of 2148 bp in length
*	35986	36085	gap of unknown length
*	36086	39468	contig of 3363 bp in length
*	39469	39568	gap of unknown length
*	39569	43008	contig of 3440 bp in length
*	43009	43108	gap of unknown length
*	43109	45505	contig of 2397 bp in length
*	45506	45605	gap of unknown length
*	45606	48876	contig of 3271 bp in length
*	48877	48976	gap of unknown length
*	48977	52243	contig of 3267 bp in length
*	52244	52343	gap of unknown length
*	52344	56160	contig of 3817 bp in length
*	56161	56360	gap of unknown length
*	56361	60475	contig of 4215 bp in length
*	60476	60575	gap of unknown length
*	60576	64722	contig of 4147 bp in length
*	64723	64822	gap of unknown length
*	64823	66816	contig of 3794 bp in length
*	66817	68716	gap of unknown length
*	68717	73739	contig of 5023 bp in length
*	73740	73946	gap of unknown length
*	73947	79446	contig of 5507 bp in length
*	79447	83302	contig of 3856 bp in length
*	83303	83402	gap of unknown length
*	83403	90225	contig of 6823 bp in length
*	90226	90325	gap of unknown length
*	90326	96171	contig of 5846 bp in length
*	96172	96277	gap of unknown length
*	96278	101663	contig of 5332 bp in length
*	101664	101763	gap of unknown length
*	101764	108594	contig of 6831 bp in length
*	108595	108694	gap of unknown length

FEATURES	source
*	108695 114283: contig of 5589 bp in length
*	114284 114383: gap of unknown length
*	114384 120900: contig of 6517 bp in length
*	120901 121000: gap of unknown length
*	121001 129340: contig of 8340 bp in length
*	129341 129440: gap of unknown length
*	129441 138422: contig of 8982 bp in length
*	138423 138522: gap of unknown length
*	138523 149684: contig of 11162 bp in length
*	149685 149784: gap of unknown length
*	149785 160420: contig of 10656 bp in length
*	160421 160520: gap of unknown length
*	160521 175120: contig of 14600 bp in length.
	location/Qualifiers
	1. .175120

BASE COUNT	a	c	g	t	others
ORIGIN	41850	41856	41556	43211	6647

Query Match	73.5%;	Score	793.8;	DB 2,	Length	175120;	
Best Local Similarity	85.4%;	Pred.	No. 1.8e-122;				
Matches 922; Conservative	0;	Mismatches	148;	Indels	10;	Gaps	3;

OY	1	ATGGCCAGCCATTCTTGTGATGAATATCAACACCAGCCTTCAAACTAGCAACTGTACAGGT	60
Db	72106	ATGGCTAGCCACTCTTGTGATGAATCAACACCAGCCTTCAAACTAGCAACTGTACAGGT	72165
OY	61	GCTGCTCTGTCCAGGAAGAGCTGAACCCCTGAGCAGCCGCCAGAGCGGAGAGAGCGAGGT	120
Db	72166	GCTGCTGTGTGCCATGAGGAGCAAGAACTCTGAGCGCCGCCCAAGAGCGGAGAGAGCGAGGT	72235
OY	121	CCCGAGGAGGACAGTAAAGTGGCAATGAGAGCGTTCCCCAGTTGGGTGCGGTCCGGGG	180
Db	72226	CCTAAGGAGGACAACTAGCTGGCAATGAGAGCGCTCTTGCACTGCGGTAGCCGTCCAGGG	72285
OY	181	CCGAGAGGGGGAAGGAGAGCTGTGAATCCCAACCACTCCCTTGGAGAGCCAGGCGTTTCCA	240
Db	72286	CAGAGGGGAGACGGGGGCGCTGAAGACACACACTGCCCCCAATGTGAGACGAATGCTGTCCA	72345
OY	241	GAATCTAGCTGCTGAGAGAGGGGCGAGAGGGGCCAATAGGGAGAGCACTGTCCGCTGGC	300
Db	72346	GAATCTAGCTGCTTGGAAAAAGGGCGAGAGGGCCAGAAATGGAGGACCTTATCCACTGGCC	72405
OY	301	GGGCACTTCCGGCGCGGCGGAGAAAGTGGAAACGCAACCGGAGGGCGAGGTGCTCGGCCAG	360
Db	72406	GGTGCCT---CGCCTTGCGGAGGGGAGGCCGATGCAGAGTCC-----CTCGTGCAA	72456
OY	361	CCTTGTATGATACCTCCGAGGCCAGTAAAGTTGGGGGCTCTGCGCGAGAGGGGCGAAGAGAG	420
Db	72457	CCAGGTGATGACTGCGGAGGCTACTAAAGCTGAGGCTCTGTGCTGCTGGAATANAGAGCCA	72518
OY	421	TGGGAGACAGACGACGAGACAGCTGGGGAAAGAAAAACATTAAGAGACGCCGCTCCAAAGAG	480
Db	72517	TGGGAGACAGACGACGAGACAGCTGGGCAAGAAAAACACCGGAGAGCCCTCTCCAGGAAG	72576
OY	481	AAGGCGCATTTGGAAACCCTACTCAAAAGCTGMACTGGGAAGAGAAAGAAAAAGTTTCAGCAG	540
Db	72577	AAGCGACATTTGGAAACCCTATTTACAAAGGTG-AC TGGGAGAGAGAAAGAAAAAGTTTCAGCAG	72635
OY	541	AAACGAGAGCCTTGAGGCTTCAAGATCCGAGCCGGAATGTGTGCCCAAAGGGCCAGCCGGTC	600
Db	72636	AAGGAAAGCCCTGGAGGCTTGCGAGTTGCGAGCCGGAATGTGTGCCCAAAGGGCCAGCCGGTT	72695
OY	601	GGCGCTATTAACACACAGCAGCTTCTCATGATGATGATCAGAGCAGGAGGAGCGGATCTC	660
Db	72696	GGCGCTATTAACACACAGCAGCTTCTCATGATGATGATCAGAGCAGGAGGAGGCTGATCTC	72755
OY	661	AAAAACGGGCTGTATCTCCAAGGGGGCGCCGCCAAATTCGAGAGCACACGCGATGACGAC	720
Db	72756	AAAAACGGGCTTATACCTTAAGCGGGGAGCGCGCCAAATTCGAGAGCACACGCGATGAGGAT	72815

BASE COUNT	GCRDEPCT*			
ORIGIN	324 a	338 c	392 g	184 t
Query Match	16.2%; Score 174.8; DB 9; Length 1238;			
Best Local Similarity	60.5%; Pred. No. 1.6e-19;			
Matches	362; Conservative 0; Mismatches 197; Indels 39; Gaps 3;			
QY	441	GCTGGGAGAGAAAAACATAAGAGACGCCCGTCCACAGAGAGCGGCGATTGGAAACCGTA	500	
DB	344	GCTGGCCCGGAGAAACACCGCTCGCGCCATCCAGACGCAAAAAGCACTGGCCACCTTA	403	
QY	501	CTACAGCTGAACAGGAGAGAAAAAGTTGACAGCAAGAACAGACCTTCGAGCTTC	560	
DB	404	CTTGAGACTGAGCTGGGCTGAGAAACAACAGCGGGATGAGAGGACCCAGAGGCGCTC	463	
QY	561	AAGATCCGAGCCGAGATGTTGGCCAAAGGCCAGCGGTCGCGCCCTATAACACACGCA	620	
DB	464	CCGGGTCGGGAGAGATGTTGCGCAAAAGGCCAGCCCGTGGCCCTACAAACACACCCA	523	
QY	621	GTTCCCTATGATGATACACACAGAGAGAGACCGGATCTCAAAACCGGCTGTACTCAA	680	
DB	524	GTTCTCTATGATGATACAGGAGACCGGAGAGACC-----CAACTTGAGATGTCGCCCAT	575	
QY	681	GGCGGCCCGCCGCAAAATCCGACAGACACGAGATGACACTTTCATGAGAAAGAGGGGTGA	740	
DB	576	GGGATCTCCACCCAGGTTCCAG-----TGGGGAGAG	607	
QY	741	GGAGATGGGGCGGCGCATGGAGTGGAGGGGACGGACGAGTTTCTGCAGCGGAGATT	800	
DB	608	TGAGGCCGGGAGACAGTATGGGCGGGCCGAGCCGACGCTGAGTTCCAGCGGAGAGACTT	667	
QY	801	CTGGAGACAGCAGAGGGGTATACACAGAGAGGCTCTGAGAACATGAGCAAGAGAGACT	860	
DB	668	CTCTGAGACTTACAGAGGCTTCCACACGAGAGGCTTCTGAGGCGCCGACAGAGAGGACT	727	
QY	861	CATCAAGAGTACTCTGAGACTGAGAGAGTGCCTCTCCGCGATGGAGAGACAGCAACACG	920	
DB	728	GGTCGCGAGCTACTGAGCTGAGAGAGCGGCTGTCCAGGCGGAGAGGAGACTTAGAG	787	
QY	921	GCTCGGCTGAGAGCAACGCGCTGGTGGCGA---CGACGCGGCTGTGCGGAGACTGGA	977	
DB	788	GCTGCACAGCTGAGGCGCTGCACCGGCGACGACTCTGCCGAGGAGGAGAGCTGGC	847	
QY	978	GCTGAGACTGGACCGCGCGCGGAGAACTTCAGCTGCTGACCGCAAGCAAGTG	1035	
DB	848	TGCCGAGGTCAGAGGCTCCGGAGCCGAAACACAGCGGCTTCTGAGAGAGACAGATG	905	
RESULT 12				
AK056946	1330 bp mRNA linear PRI 01-AUG-2002			
LOCUS	Homo sapiens cDNA FLJ32384 f1s, clone SKMUS1000104, weakly similar			
DEFINITION	to Homo sapiens mRNA for HEXIM1 protein, complete cds.			
ACCESSION	AK056946			
VERSION	AK056946.1 GI:16552484			
KEYWORDS	oligo capping; f1s (full insert sequence).			
SOURCE	Homo sapiens skeletal muscle cDNA to mRNA, clone_lib:SKMUS1 clone:SKMUS1000104.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS	Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K., Arita,M., Musashino,K., Youki,H., Hara,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai,Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Magatsune,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuko,Y., Nagai,K. and Isogai,T.			
TITLE	NEO human cDNA sequencing project			
JOURNAL	unpublished			

REFERENCE	2 (bases 1 to 1330)			
AUTHORS	Isogai,T., Otsuki,T. and Sugiyama,T.			
JOURNAL	Direct Submission			
COMMENT	Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-338-52-3975, Fax:81-438-52-3986) NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.			
FEATURES	Location/Qualifiers			
Source	1..1330			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="SKMUS1000104"			
	/tissue_type="skeletal muscle"			
	/clone_lib="SKMUS1"			
	/note="Cloning vector: pME18SFL3"			
	240..1100			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="BAB71319.1"			
	/db_xref="GI:16552485"			
	/translation="MMATPNOTACNAESPVALAEKATSGAPSGFQTPPEPRHDSGSLP LTPRMESHSEDEDLAVGAVGGVGNMSPRQSPGCAEAVLARKHRRPSSKRRHM RPYLELSMAKQOQDEROSORASHREMEAKGQPVAPYNTQFLMDRPPSEKRLNDV PHGISHPGSGSEFAGDSGDSRAHGFORRDFEYERFHTESLQGRSROELVADYL ELERLSQAEETFRRLQQLACTQSQGRQVEELAAVORLRTENQLRQENQMWNE GCRDEPCT"			
BASE COUNT	295 a	370 c	452 g	213 t
ORIGIN				
Query Match	16.2%; Score 174.8; DB 9; Length 1330;			
Best Local Similarity	60.5%; Pred. No. 1.6e-19;			
Matches	362; Conservative 0; Mismatches 197; Indels 39; Gaps 3;			
QY	441	GCTGGGAGAGAAAAACATAAGAGACGCCCGTCCACAGAGAGCGGCGATTGGAAACCGTA	500	
DB	494	GCTGGCCCGGAGAAACACCGCTCGCGCCATCCAGACGCAAAAAGGCACTGGCCACCTTA	553	
QY	501	CTACAGCTGAACAGGAGAGAAAAAGTTGACAGCAAGAACAGAGCCTTCGAGCTTC	560	
DB	554	CTTGAGACTGAGCTGGGCTGAGAAACAACAGCGGGATGAGAGGACAGAGGCGCTC	613	
QY	561	AAGATCCGAGCCGAGATGTTCCGCAAGGCGCACCGGTCGCGCCCTATAACACACGCA	620	
DB	614	CCGGGTCGGGAGAGATGTTGCGCAAAAGGCCACCGCTGGCCCTACAAACACACCCA	673	
QY	621	GTTCCCTATGATGATACACACAGAGAGAGCGGATCTCAAAACCGGCTGTACTCAA	680	
DB	674	GTTCTCTATGATGATACAGGAGACCGGAGAGACC-----CAACTTGAGTGTGCCCAT	725	
QY	681	GGCGGCCCGCCGCAAAATCCGACAGACACCGAGATGAGAGACTTCTGAGAAAGAGGGGTGA	740	
DB	726	GGGATCTCCACCCAGGTTCCAG-----TGGGGAGAG	757	
QY	741	GGAGATGGGGCGGCGCATGGAGTGGAGGGGACGCGACGAGTTTCTGCAGCGGAGATT	800	
DB	758	TGAGGCCGGGAGAGATGATGGGGGGGCGGAGCCACGATGATCCAGGAGAGGACTT	817	
QY	801	CTCGAGAGCTGAGAGCGGATACACACAGAGAGCGCTTCAGACAATGAGCAAGAGACT	860	
DB	818	CTCTGAGACTTACGAAGGCTTCCACACGAGAGCCTCAGAGGCGGCAAGCGAGGACT	877	
QY	861	CATCAAGAGTACTGAGACTGAGAGAGTGCCTTCGCCCATGAGGAGACGAGAACACCG	920	
DB	878	GTCGCGAGACTTACCTGGAGTGGAGAGCGGCTCTCGAGGCGGAGAGGAGACTTAGAG	937	
QY	921	GCTCGGCTGAGAGCAAGCGGCTGGGTGGCGA---CGACCGCGCTGTGCGGAGACTGGA	977	

Db 938 GCTGACAGCTGCAGGGCGTGCACCAGCCGACAGACTCCTTCGCCGCCAGGTGGAGAGAACTGGC 997
|||||+||++||-|-|+||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||
Dy 978 GCTGAGCTGTGACCGCGCTGCAGCGCCGAGAACCTTCACCTGCTGTGAACCGAACAAGCACTG 1035
||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||-||
Db 998 TGCCGAGGATCCAGAGGCTCCGAGCACGAAAACACAGCGGCTTGCTGACGAGAACAACGATTG 1055
+++++-----+-----+-----+-----+-----+-----+-----+-----+

RESULT 13

LOCUS AC008105

DEFINITION Homo sapiens chromosome 17 clone CTD-2020K17 map 17, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.

ACCESSION AC008105

KEYWORDS HTGS_105_24 GI:21617753

SOURCE HTGS_PHASEI; HTGS_FULLTOP; HTGS_ACTIVEFIN.
human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 116035)
Britten,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone CTD-2020K17
Unpublished

AUTHORS 2 (bases 1 to 116035)
Britten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArrellano,K., Depierre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,K., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Headford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatsas,A., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., Mclaughlin,J.,
Mejdum,J., Molina,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,T., Nijhoff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Sudramanian,A.,
Tesfaye,S., Torturella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

JOURNAL Direct Submission

TITLE Submitted (22-JUN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 116035)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouknighter,B.,
Cammarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham.L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson.R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Ton,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,V.,
McCarthy,M., Mejdum,J., Meneses,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Notcu,C., Norman,C.H.,
O'Connor.T., O'Donnell.P., O'Neill.D., Oliver.J., Peterson.K.,
Phunkhang.P., Pierre.N., Raymond.C., Retta,R., Rise.C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schnuppach,R., Seman,S., Sevely,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talmas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers.M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu.X., Wyman,D., Young.G., Zainoun,D.,
Zembek,L., Zimmer.A. and Zody.M.

TITLE Direct Submission

JOURNAL Submitted (22-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jun 27, 2002 this sequence version replaced gi:20148201.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L890
Center clone name: 2020_K_17

* NOTE: THIS is a working draft sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number WILL
* be preserved. 6930. *GenBank* 16930. *GenBank* 16930.

```
*
* 1      16830: contig of 16830 bp in length
* 16831 16930: gap of 100 bp
* 16931 29738: contig of 12808 bp in length
* 29739 29838: gap of 100 bp
* 29839 39151: contig of 9313 bp in length
* 39152 39251: gap of 100 bp
* 39252 445913: contig of 6662 bp in length
* 445914 46013: gap of 100 bp
* 46014 116035: contig of 70022 bp in length
*
```

FEATURES
SOURCE

	/clone_lib="CITD1 Human BAC"									
BASE COUNT	26107	a	30142	c	30260	g	29051	t	475	others
ORIGIN										

Query Match	16.2%	Score 174.8:	DB 2:	Length 116035:
Best Local Similarity	60.5%	Pred. No. 8.7e-20:		
Matches 362; Conservative	0;	Mismatches 197;	Indels 39;	Gaps 3;

QY	441	GCCTGGGGAAGAAAAACATAGAGACGCCCGTCCAGAGAAGACGGCATTTGGAACCCGTA	500
Db	17959	GCCTGGCCCCGAGAGAACACCGCTCGCGGCGCATCGAAGCGCCAAAAGGCACTGGCCACCCCTA	18010

0y 501 CTACAGCTGACCTGGGAGAGAGAAAAAGTTGACGAGAAACAGAGCCTTCAGCTTC 560
 18019 CCTGGAGCTGAGCTGGGCTGAGAGAAACACAGCGGGATGAGAGCGACAGAGCCAGAGGGCTTC 18077
 Db

[illegible]

Qy	621	GTTCCTCATGATGATCATCAGCACAGGAGAGAGCCCGCATCTCMAAACCGGCGCTGTGTCCCA	680
Db	18139	GTTCCTCATGATGATCATCAGGAGACCCCGGAGAGACC-----CAACTTGATGTGCCCCAT	18191

0y GCGGGGCCCCGCCAAATCCGACGCAGCACCAGCATTCATGTGAAGAAGGGGGCTA 740
||| |
681 ||| |
Db 18191 GGGA TCTCC CACCAG GTTCC AG-----TGGGGAGAG 18222

Db

OY 741 GGGAGGATGGGGGCCAGCGATGGGATGGGAGGGGACGGCAGCGATTTCGCACGGGGACTT 800
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18223 TGAGGCCGGGGACAGTGATGGGGGGGGCCGAGCCACGGTTGATGTTCACGCGGAAGGACTT 18288

[illegible]

QY 861 CATCAAGCGAGTACTTGGAACTGGAGAACTGCCCTCTCGGCGCATGGAGGACGAAACCAACG 920
 18343 GGTGCGAGACTACTTGGAACTGGAGAAAGCGGCTTTCGACGCGCGAGAGAGAACTAGGAG 18400

921 GGTGGCGGCTGGAGAGCAAGCGGCTGGTGGCGA---CGACGCGGCTGTGCGGGAGCTGGA 977
 18403 GCTTCAGCAGCTCAGGCGCTGCACCGCGCCAGCACTCTTCCTCCCGCCAGGTGGAGAGCAGCTGCC 1846

QY	978	GCTGAGCTGAGACCGCGCTCGCGCGCGGACCAACCTTCAGCTGCTGACCGGAAAGCAACTG	1035
Db	18463	TGCCGAGGTCCAGAGCGCTCGGAGCAACCAACCGAGCGCTTCCTGACGAGAACCAAGATG	18520
RESULT 14			
AK096785/c			
LOCUS	AK096785	2048 bp	mRNA
DEFINITION	Homo sapiens cDNA FLJ39466 fis, clone PROST2012353.		linear
ACCESSION	AK096785		
VERSION	AK096785.1	GI:21756354	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens prostate cDNA to mRNA, clone_lib:PROST2 clone:PROST2012353.		
ORGANISM	Homo sapiens		
REFERENCE	Eukariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,K., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Nagatsuna,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.,		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2048)		
AUTHORS	Isogai,T. and Yamamoto,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 259-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.		
COMMENT			
FEATURES			
SOURCE	location/qualifiers		
	1..2048		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="PROST2012353"		
	/issue_type="prostate"		
	/clone_lib="PROST2"		
	/note="Cloning vector: PME18SFL3"		
BASE COUNT	383 a 655 c 558 g 452 t		
ORIGIN			
Query Match	16.0%: Score 173.2; DB 9; Length 2048;		
Best Local Similarity	60.4%: Pred. No. 2.7e-19;		
Matches	361; Conservative 0; Mismatches 198; Indels 39; Gaps 3;		
QY	441	GCTGGGGAGAAAAAATTAAGACAGCCCGCTCCAGAAAGAGCGCATTTGGAACCGTA	500
Db	1650	GCTGGCCCGGAAGAAACACCGTGGCGGCATTCAGAGCGCAAAAGGCACTGGCACCTA	1591
QY	501	CTACAGCTGACTGGGAGAGAGAAAAAGTTGCGAGCAGAGAAAGAGCCCTTCAGAGCTTC	560
Db	1590	CCTGGAGCTGAGTGGCGCTGAGAAACAACACGCGGATGAGAGGACCGAGAGGCGCTC	1531
QY	561	AAGATATCCGAGCCGAGATGTTCCGCCAAGAGGCGCAGCCGCTTAAACCGCGCTGTACTCCA	620
Db	1530	CGGGGTCGGCAAGAGATGTTGCGCAAAAGGCGACCCGCTGGCCCCCAACAACACCA	1471
QY	621	GTTCCTCATGATGATCATCAGCAGAGAGAGCGCGATCTCAAACCGCGCTGTACTCCA	680
Db	1470	GTTCCTCATGATGATCATCAGAGAGACCGGAGAGAGCC-----CAACTTGGATGTGCCCCAT	1419
QY	681	GCGGCGCGCGCAAAATCCGACGACACCGAGCATGACGACTTCAATGGAAGAGGGGGTGA	740

Db	1418	GGGATCTCCACCCAGGTTCCAG-----TGGGGAGAG	1387
OY	741	GGAGATGATGGGACAGCATGGGATGGGAGGACGACGAGTGTTCGACGCGGACTT	800
Db	1386	TGAGGCCGGGACACAGTATGGGCGGGGCCGACGCGATGATGTTCCAGCGGAAGACTT	1327
OY	801	CTCGAGACGTACGAGCGGTATCCACACGAGGAGCGCTGCAGAACATGAGCAACAGCAGACT	860
Db	1326	CTTGAGACTTACGAAACGCTTCCACACGAGAGGCTTGAGGAGCGCCGACGAAACAGGAGCT	1267
OY	861	CATGAAGGAGTACTTGGAACTGGGAAAGTGCCTCTCGCGCATGGAGACGAAACAAACCG	920
Db	1266	GGTCCGAGACTACTCTGAGCTGGAGAAAGCGGCTCTCCAGGCGAGGAGGAGACTTAGAG	1207
OY	921	GCTCGCGTGTGAGAGCAAGCGGCTGGGGGCGCA---CGACGCGGCTGTGGCGAGCTGGA	977
Db	1206	GCTCAGCAGCTGCAGAGCGGTGCACCGGCGACGACTCTCGCGCAGAGTGGAGGAGCTGGC	1147
OY	978	GCTGAGCTGAGACCGGCTGCAGCGCCGAGAACCTTCAGCTGCTGACCGAGAACGAACTG	1035
Db	1146	TGCCGAGGTCCAGAGGCTCCGAGCCGAAACACAGCGGTTGTGTCAGAGAAACGAGATG	1089
RESULT 15			
AL662804	223442 bp	DNA	linear
LOCUS			ROD 26-JUN-2002
DEFINITION	Mouse DNA sequence from clone RP23-341C5 on chromosome 11, complete		
ACCESSION	AL662804		
VERSION	AL662804.17		
KEYWORDS	GI:21621620		
SOURCE	HTG.		
ORGANISM	house mouse.		
	Mus musculus.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 223442)		
AUTHORS	Blakey, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk		
COMMENT	On Jun 27, 2002 this sequence version replaced gi:21615602.		

	Center: Wellcome Trust Sanger Institute		
	Web site: http://www.sanger.ac.uk		
	Contact: humquery@sanger.ac.uk		

	During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.		
	This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP. Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-341C5 is from the RP23-Mouse PAC Library constructed by the group of Plier de Jong. For further details see http://www.chori.org/bacpac/home.htm		
	VECTOR: pBAC3.6.		
	Location/Qualifiers		
	1..223442		

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-341C5"
/clone_lib="PCL1-23"
BASE COUNT      56058 a 54948 c 56347 g 56089 t
ORIGIN

```

```
/db_xref="taxon:10090"
```

```

/chromosome="11
/chromosome="PP33=3A

```

```
/clone- lib="RPCT-23"
```

54948 c 56347 q

ORIGIN

ORIGIN

Query Match	Score	DB	Length
15.98;	172;	10;	223442;

Best Local Similarity	59.28;	Pred. No. 2.3e-19;
Matches	354;	Conservative
0;	Mismatches	205;
Indel	39;	Gap
2;		

machines 227, conservative 0, minimums 203, insects 23, vaps 24

QY	441	GCTGGGGAAGAAAAACATTAAGAACGGCCCTCCCAAGAAAGGCGGCTTTGGAAACCGTA	500
Db	3891	GCTGGCCCGAAAAGAACCCCTAAGCGGCCCTTCAGAGCGCAAGGGGCACTGGCGCCCTA	3950
QY	501	CTCAAGCTGTAACGTGGGAAGAAAGAAAAGTTTCGACGAGAAACAGACCCCTTCGAGCTTC	560
Db	3951	CTTGAGAGTGAGTGGGGCCGGAAGAACACGAGAGATGAGAGGCAAGCCAGAGGGCCCTC	4010
QY	561	AAGATCCGAGCCGAGATGTTGGCCAAAGGCCAGCCGGTCCGGCCCTATACACCACGCA	620
Db	4011	TCGGGTCCCGTAGAGAGATGTTCCGCAAAAGGTACACCCCTCGCGCCCTACCAACACACC	4070
QY	621	GTTCCCTATGATGATATACAGCAACGAGAGAGATCCGGATCTCAAAAACGGCCCTGTA	680
Db	4071	GTTCCCTATGATGATATGATGCTGACCTCGAGAGACCTTAACCTCGATGCTTCAACGG	4130
QY	681	GCGGGCCGCCCAATTCGACGACACCGCATGACGACTTCATGGAAGAAAGGGGTGA	740
Db	4131	CCACTCTGGGCTC-----GGTGGGAGAA	4154
QY	741	GGAGGATGGGGGACCGCATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCACGGGACTT	800
Db	4155	CGAAGCAGGGGACAGTGTGGGCAAGGGCCGAGCCCATGGGGAAATTTACGACAGGACTT	4214
QY	801	CTCGAGACGTACACGGGTACCAACACGGAAGCCTGCAGAAACATGACCAAGCAGGAGCT	860
Db	4215	CTCTGAGGCCCTATATAGCGGTACCACTCAGAGGCTTTAGGGGCGGCACCAACGACGAGCT	4274
QY	861	CATCAAGGAGTACCTGGAACCTGGAGAGTGCCTCTCGGCATGTGAGGAGACGAAACACG	920
Db	4275	GGTCGAGACTACCTGGATATGAGAGGGCGTCTGTACAGAGCTGAGCAAGAAACTCGGAG	4334
QY	921	GCT---GCGGCTGGAGAGCAAGCGGCTGGGTGGCGACAGCGCGCTGTGCGGAGCTGGA	977
Db	4335	GCTCCGGAGCTGACAGGGTTGCTCCAGCAGGCAACCCGTCTCAACAGAGTGAAGAGCTGGC	4394
QY	978	GCTGAGGTGAGCGGCTGGCGGCGGAGAACTCCAGCTGCTGACGAGCAAGCAAGT	1035
Db	4395	TGCTGAGGTGAGAGGCTCCACAACTGAAACACGAGCGGTGGCTGAGGAGAAACGAGATG	4452

Db 3891 GCTGCCCGAAGAGCACCGTAGGGCGCCATCGAAGCGCAAGCGGCACTGGCGGCCCTA 3950

3951 CTTGAGCTGAGCTGGGCCGAGAAGCAACAGCGAGATGAGAGGGCAGAGCCAGAGGGCCTC 4010

Db 4011 TCGGGTCCGTGAGAGATGTCGCCAAGGTCAGCCCCCTGGCCCTTACACACACCCCA 4070

Db 4071 GTTCCCATGATGATGTCGTGACCTGGAGGAGCCCTAACCTCGATGTGCTTCACGGGGCCCTC 4130

Db 4131 CCACCTCTGGCTCC-----GGTGGGGAGCAA 4154

Db 4155 CGAAGCAGGGGACAGTGAATGGGCAAGGCCGAGCCCATGGGSAATTTTCAGCAGAGGGACTT 4214

Db 4215 CTCTGAGGGCTTATGAGCGGTACACACTTACAGAGCCTTCAGGGCCGACGACAGGAGACT 4274

Db

4275 GGTGGCAGACTTACCTCGATCTTAGAGAGGGCTCTGTGCACAGGCTGACGACGGAACCTCGAG 4334

4335 GCTCCGGCAGCTGCAGGCTTGCTTCACGACAGGCAACCCCTGTCACAGCTGGAAAGAGCTGGC 4394

Db 4395 TGCTGAGGTGGAGAGGCTCAGAACTGAATACCACAGCGGCTGCTCAGGAGAACGAGATG 4452

Search completed: June 18, 2003, 01:05:33
Job time : 2958 secs

Search completed: June 18, 2003, 01:05:33
Job time : 2958 secs

Job time : 2958 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 01:11:07 : Search time 949 Seconds
(Without alignments)
1669.991 Million cell updates/sec

Title: US-09-972-758a-1
Perfect score: 1080
Sequence: 1 atgacgcagacattcttcttc.....tttccaaagtttgagactag 1080

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications_NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PTCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1080	100.0	1080	9 US-09-972-758-1	Sequence 1, Appli
2	1076.8	99.7	2199	9 US-09-745-763-33	Sequence 33, Appl
3	1076.8	99.7	3624	9 US-10-205-823-171	Sequence 171, App
4	368.8	34.1	495	9 US-09-918-995-403	Sequence 403, App
5	362.8	33.6	414	10 US-09-983-965-4895	Sequence 4895, App
6	355.2	32.9	461	9 US-09-954-531-813	Sequence 531, App
7	68.4	6.3	8705	9 US-10-291-230-14	Sequence 14, Appl
8	67.8	6.3	4270	9 US-09-954-531-1352	Sequence 1352, Ap
9	67.8	6.3	4270	10 US-09-880-107-2427	Sequence 2427, Ap
10	60.8	5.6	2108	10 US-09-962-832-225	Sequence 225, App
11	58.4	5.4	2313	9 US-09-854-133-157	Sequence 157, App
12	58.4	5.4	2313	10 US-09-738-973-157	Sequence 157, App
13	58.4	5.4	2314	9 US-09-764-868-12	Sequence 12, Appl
14	53.6	5.0	607	9 US-10-043-487-122	Sequence 122, App
15	53.6	5.0	1609	9 US-10-043-487-105	Sequence 105, App
16	52	4.8	225883	9 US-10-175-523-57	Sequence 57, Appl
17	50.2	4.6	1949	9 US-10-274-968-1	Sequence 1, Appli
18	50.2	4.6	1949	10 US-09-740-027-1	Sequence 1, Appli
19	49.6	4.6	506	9 US-10-060-036-222	Sequence 222, App

20	49.6	4.6	1327	9 US-10-043-487-168	Sequence 168, App
21	49.4	4.6	594	9 US-10-123-155-10	Sequence 10, Appl
22	49.2	4.6	14800	10 US-09-954-456-1601	Sequence 1601, Ap
23	49	4.5	3951	9 US-10-029-115-1	Sequence 1, Appli
24	49	4.5	4414	9 US-10-029-115-3	Sequence 3, Appli
25	48.2	4.5	401	9 US-09-918-995-37237	Sequence 37237, A
26	48.2	4.5	549	9 US-09-991-936-1712	Sequence 1712, Ap
27	47.4	4.4	1101	10 US-09-874-062-2	Sequence 2, Appli
28	47.4	4.4	1482	9 US-09-789-390-64	Sequence 64, Appl
29	47.4	4.4	1649	9 US-09-789-390-62	Sequence 62, Appl
30	47.4	4.4	1717	9 US-09-789-390-69	Sequence 69, Appl
31	47.4	4.4	1787	9 US-09-789-390-66	Sequence 66, Appl
32	47.4	4.4	3735	9 US-09-789-390-12	Sequence 12, Appl
33	47.4	4.4	3822	9 US-09-789-390-10	Sequence 10, Appl
34	47.4	4.4	3912	9 US-09-789-390-8	Sequence 8, Appli
35	47.4	4.4	3999	9 US-09-789-390-6	Sequence 6, Appli
36	47.4	4.4	4033	9 US-10-029-115-5	Sequence 5, Appli
37	47.4	4.4	4133	9 US-09-291-417-11	Sequence 11, Appl
38	47.2	4.4	520	9 US-10-184-644-332	Sequence 332, App
39	47.2	4.4	520	9 US-10-184-634-332	Sequence 332, App
40	47	4.4	3511	10 US-09-753-831-3	Sequence 3, Appli
41	47	4.4	4184	9 US-09-927-827-42	Sequence 42, Appl
42	46.8	4.3	971	9 US-10-102-806-61	Sequence 61, Appl
43	46.8	4.3	2411	10 US-09-823-356-23	Sequence 23, Appl
44	46.8	4.3	3259	12 US-10-001-843-71	Sequence 71, Appl
45	46.8	4.3	24707	9 US-10-274-968-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-972-758-1
Sequence 1, Application US/09972758
Patent No. US20020160497A1
GENERAL INFORMATION:
APPLICANT: Case Western Reserve University
APPLICANT: Montano, Monica
APPLICANT: Miltman, Bryan
TITLE OR INVENTION: Suppressors of Human Breast Cancer Cell Growth
FILE REFERENCE: 27708/04004
CURRENT APPLICATION NUMBER: US/09/972,758
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: US 60/238,187
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 1080
TYPE: DNA
ORGANISM: Homo sapiens
US-09-972-758-1

Query Match 100.0%; Score 1080; DB 9; Length 1080;
Best Local Similarity 100.0%; Pred. No. 2e-278;
Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGCCGAGCATTCTTGTGAGATATCAACACCGCTTAACACTAGCACTACAGCT	60
DB	1	ATGGCCGAGCATTCTTGTGAGATATCAACACCGCTTAACACTAGCACTACAGCT	60
QY	61	GGTGTCTCTTCCAGGAAGAGCTGAACCCCTGACGCCCCCAGCGGGGAGGGGCTG	120
DB	61	GGTGTCTCTTCCAGGAAGAGCTGAACCCCTGACGCCCCCAGCGGGGAGGGGCTG	120
QY	121	CCGAGAGGAGACAGTAGTGAGCAATCGAGAGCTTCCCGCAGTTGGGTGCGCTCGGG	180
DB	121	CCGAGAGGAGACAGTAGTGAGCAATCGAGAGCTTCCCGCAGTTGGGTGCGCTCGGG	180
QY	181	CCGAGAGGAGAGGAGAGCTTGAATCCCAACACCTCTTGGAGAGCCAGGCTGTCCA	240
DB	181	CCGAGAGGAGAGGAGAGCTTGAATCCCAACACCTCTTGGAGAGCCAGGCTGTCCA	240

241 GAATCTAGCTGCTGAGAGAGAGGCGAAGAGGCGCAGAACTGGGAGCAGCTGTCCTGCGTGC 300
241 GAATCTAGCTGCTGAGAGAGGCGAAGAGGCGCAGAACTGGGAGCAGCTGTCCTGCGTGC 300
301 GGGGACTTCCCGCCCGCCGAGAGAGTGGAAAGCCAGAGCCCGAGAGCCGAGCTGCTGCGCCAG 360
301 GGGGACTTCCCGCCCGCCGAGAGTGGAAAGCCAGAGCCCGAGAGCCGAGCTGCTGCGCCAG 360
361 CCTGTGATGACTCCGAGAGCCAGTAAGTTGGGGGCTCTGCGCAGAGGGGGCGCAAGAGAG 420
361 CCTGTGATGACTCCGAGAGCCAGTAAGTTGGGGGCTCTGCGCAGAGGGGGCGCAAGAGAG 420
421 TGGGAGACAGCAGAGAGAGAGCTGGGGAAGAAAAACATTAAGAGCCCGCTGCAAGAAG 480
421 TGGGAGACAGCAGAGAGAGAGCTGGGGAAGAAAAACATTAAGAGCCCGCTGCAAGAAG 480
481 AAGCGCATTTGGAAACCGTACTACAAAGCTGAAGTGGGAAGAGAAAGTTTCGAGAG 540
481 AAGCGCATTTGGAAACCGTACTACAAAGCTGAAGTGGGAAGAGAAAGTTTCGAGAG 540
541 AAACAGAGCCCTTGAAGCTTCAAGAGATCCGAGCCGAGATGTTGCGCAAGGGCCAGCGGTC 600
541 AAACAGAGCCCTTGAAGCTTCAAGAGATCCGAGCCGAGATGTTGCGCAAGGGCCAGCGGTC 600
601 GGGCCCTATACACACAGCAGTCTCTCATGATGATCAGCAGCAGAGAGAGCGGATCTC 660
601 GGGCCCTATACACACAGCAGTCTCTCATGATGATCAGCAGCAGAGAGAGCGGATCTC 660
661 AAAACCGGCTTACTCCAAAGCGGGCGCGCCCAATCCGAGAGACCCAGCGATGCGAGC 720
661 AAAACCGGCTTACTCCAAAGCGGGCGCGCCCAATCCGAGAGACCCAGCGATGCGAGC 720
721 TTCATGAGAAAGGGGGGTGAGAGAGATGGGGGCGAGCGATGGGAGGGGAGCGGAGCAG 780
721 TTCATGAGAAAGGGGGGTGAGAGAGATGGGGGCGAGCGATGGGAGGGGAGCGGAGCAG 780
781 GAGTTTCGACGCGGAGCTTCTCGAGAGAGTACGAGCGGTACCACACGAGAGAGCTGCGAG 840
781 GAGTTTCGACGCGGAGCTTCTCGAGAGAGTACGAGCGGTACCACACGAGAGAGCTGCGAG 840
841 AACATGAGCAACAGAGAGCTCATCAAGAGTACTGGAAGTCTGGAAGTCTGCGGCGC 900
841 AACATGAGCAACAGAGAGCTCATCAAGAGTACTGGAAGTCTGGAAGTCTGCGGCGC 900
901 ATGGAGAGAGAGAGAGAGCGGCTGCGGCTGAGAGAGAGCGGCTGGGTGGCGAGAGCGG 960
901 ATGGAGAGAGAGAGAGAGCGGCTGCGGCTGAGAGAGAGCGGCTGGGTGGCGAGAGCGG 960
961 CGTGGCGGAGCTGGAGCTGGAGTGGAGCGAGCGGCTGGCGGCGGAGAACTCCAGCTGCTG 1020
961 CGTGGCGGAGCTGGAGCTGGAGTGGAGCGAGCGGCTGGCGGCGGAGAACTCCAGCTGCTG 1020
1021 ACCGAGAAAGAACTGACCGGAGAGAGAGAGCGGCGCTTCCAAAGTTGGAGAGTAG 1080
1021 ACCGAGAAAGAACTGACCGGAGAGAGAGAGCGGCGCTTCCAAAGTTGGAGAGTAG 1080

RESULT 2
US-09-745-763-33
Sequence 33, Application US/09745763
Patent No. US20020065394A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallee, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merberg, David
Tresay, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 2199 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: <Unknown>
US-09-745-763-33
Query Match 99.7%; Score 1076.8; DB 10; Length 2199;
Best Local Similarity 99.8%; Pred. No. 1.7e-277;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 ATGCCGAGACCACTTTCTTGTGAGATATCAACACCAAGCTCAAACTAGCACTGACAGT 60
707 ATGCCGAGACCACTTTCTTGTGAGATATCAACACCAAGCTCAAACTAGCACTGACAGT 766
61 GCTGCTGCTGTCAGAGAGAGTGAACCTGAGCGCCCGCCGAGCGGAGAGAGCGGGTG 120
767 GCTGCTGCTGTCAGAGAGAGTGAACCTGAGCGCCCGCCGAGCGGAGAGAGCGGGTG 826
121 CCGGAGAGAGAGAGAGTGAACCTGAGCGCCCGCCGAGCGGAGAGAGCGGGTG 180
827 CCGGAGAGAGAGAGTGAACCTGAGCGCCCGCCGAGCGGAGAGAGAGCGGGTG 886
181 CCGGAGAGAGAGAGTGAACCTGAGCGCCCGCCGAGCGGAGAGAGAGCGGGTG 240
887 CCGGAGAGAGAGAGTGAACCTGAGCGCCCGCCGAGCGGAGAGAGAGCGGGTG 946
241 GAATCTAGCTGCTGAG 300
947 GAATCTAGCTGCTGAG 1006
301 GGGGACTTCCCGCCCGCCGAGAGAGTGGAAAGCCAGAGCCCGAGAGCTGCTGCGCCAG 360
1007 GGGGACTTCCCGCCCGCCGAGAGAGTGGAAAGCCAGAGCCCGAGAGCTGCTGCGCCAG 1066
361 CCTTGTATGACTCCGAGAGCCAGTAAAGTTGGGGGCTCTGCGCCGAGAGAGAGAGAGAG 420
1067 CCTTGTATGACTCCGAGAGCCAGTAAAGTTGGGGGCTCTGCGCCGAGAGAGAGAGAGAG 1126
421 TGGGAGACAGCAG 480
1127 TGGGAGACAGCAG 1186
481 AAGCGCATTTGGAAACCGTACTACAAAGCTGAAGTGGGAAGAGAGAGAGAGAGAGAGAGAG 540
1187 AAGCGCATTTGGAAACCGTACTACAAAGCTGAAGTGGGAAGAGAGAGAGAGAGAGAGAGAG 1246
541 AAACAGAGCCCTTGAAGCTTCAAGAGATCCGAGCCGAGATGTTGCGCAAGGGCCAGCGGTC 600

|||||
Db 1247 AACAGAGCCTTCGAGCTTCAGAGATCCGAGCCGAGATGTCGCCAAGGGCCAGCCGGTC 1306
QY 601 GCGCCCTATACACACGAGAGTCTCTCATGATGATCAGACAGGAGAGCCGGATCTC 660
Db 1307 GGGCCCTATMACACACGAGTCTCTCATGATGATCAGACAGGAGAGCCGGATCTC 1366
QY 661 AAAACCGGCTGTACTCCAAAGGGGGGGGGCCGCAATCCGAGACACCGAGCATGACGAC 720
Db 1367 AAAACCGGCTGTACTCCAAAGGGGGGGGGCCGCAATCCGAGACACCGAGCATGACGAC 1426
QY 721 TTCATGGAAGAGGGGGGTGAGAGGATGGGGGACGATGGGATGGAGGGAGCGGACGAC 780
Db 1421 TTCATGGAAGAGGGGGGTGAGAGGATGGGGGACGATGGGATGGAGGGAGCGGACGAC 1486
QY 781 GAGTTTCTGACGCGGACTTCTCGAGACGTACGAGCGGTACACACGAGACGACCTGCAG 840
Db 1487 GAGTTTCTGACGCGGACTTCTCGAGACGTACGAGCGGTACACACGAGAGCCTGCAG 1546
QY 841 AACATGAGCAGCAGAGGCTCATCAAGAGTACCTGGAACCTGGAGAACTGGCTTCGCGGC 900
Db 1547 AACATGAGCAGCAGAGGCTCATCAAGAGTACCTGGAACCTGGAGAACTGGCTTCGCGGC 1606
QY 901 ATGAGAGAGCAGAACACCGGCTGCGGCTGAGAGACCAACGCGCTGGGTGGCAGACGCGC 960
Db 1607 ATGAGAGAGCAGAACACCGGCTGCGGCTGAGAGACCAACGCGCTGGGTGGCAGACGCGC 1666
QY 961 CGTGTGCGGAGCTGAGCTGAGCTGAGCGCGCTGCGCGCAGACACTCCAGCTGCTG 1020
Db 1667 CGTGTGCGGAGCTGAGCTGAGCTGAGCGCGCTGCGCGCAGACACTCCAGCTGCTG 1726
QY 1021 ACCGGAACGAACTGACCGGCGGAGCAGGAGCGAGCGCGCTTCCAACTTTGGAGACTG 1080
Db 1727 ACCGGAACGAACTGACCGGCGGAGCAGGAGCGAGCGCGCTTCCAACTTTGGAGACTG 1786

RESULT 3
US-10-205-823-171
; Sequence 171, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Gorbatcheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamalkar, Shubhangi
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatz, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-044
; FILE REFERENCE: THERAPY OF PROSTATE CANCER
; CURRENT APPLICATION NUMBER: US/10/205, 823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307, 982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314, 356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325, 020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341, 746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362, 158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 171
; LENGTH: 3624
; TYPE: DNA

; ORGANISM: Homo sapiens
US-10-205-823-171
Query Match 99.7%; Score 1076.8; DB 9; Length 3624;
Best Local Similarity 99.8%; Pred. No. 1.9e-277;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 ATGGCCGAGCCATTCCTGTGACATATATCAACCCAGCCCTCAAACTAGCAACTGTACAGGT 60
Db 690 ATGGCCGAGCCATTCCTGTGACATATATCAACCCAGCCCTCAAACTAGCAACTGTACAGGT 749
QY 61 GCTGCTGCTGTCCAGAAAGAGCTGAACCTTGAGCGCCGCCAGCGCGGAGAGCGGGTGC 120
Db 750 GCTGCTGCTGTCCAGAAAGAGCTGAACCTTGAGCGCCGCCAGCGCGGAGAGCGGGTGC 809
QY 121 CCCGAGAGAGACATGATGCTGCAATCGAGAGCTTCCCACTGGTGGTGGCTCCGGGG 180
Db 810 CCCGAGAGAGACATGATGCTGCAATCGAGAGCTTCCCACTGGTGGTGGCTCCGGGG 869
QY 181 CCGGAGGGGGAGAGGAGCTTGAATCCCAACCACTCCCTTGACAGACCCAGGCTGTGCA 240
Db 870 CCGGAGGGGGAGAGGAGCTTGAATCCCAACCACTCCCTTGACAGACCCAGGCTGTGCA 929
QY 241 GAATCTAGCTGCTGAGAGAGGCGCAGAAAGGCGCAAAATGGGAGACTGTCGCTGCGC 300
Db 930 GAATCTAGCTGCTGAGAGAGGCGCAGAAAGGCGCAAAATGGGAGACTGTCGCTGCGC 989
QY 301 GCGGACTTCCGCGCGCGCAGAGATGAGAACCGACGCCCGAGCGGAGCTGCTCGCCAG 360
Db 990 GCGGACTTCCGCGCGCGCAGAGATGAGAACCGACGCCCGAGCGGAGCTGCTCGCCAG 1049
QY 361 CCTGTGATGACATCCGAGGCGCAGTAAGTTGGGGGCTCCTGCGCGAGGGGGCGAAGAGAG 420
Db 1050 CCTGTGATGACATCCGAGGCGCAGTAAGTTGGGGGCTCCTGCGCGAGGGGGCGAAGAGAG 1109
QY 421 TGGGGACACAGCAGACAGAGCTGGGGAGAAAGAAACATAGAGACGCCCTGCCAAGAG 480
Db 1110 TGGGGACACAGCAGACAGAGCTGGGGAGAAAGAAACATAGAGACGCCCTGCCAAGAG 1169
QY 481 AAGCGCATTTGGAACCCGTACTACAACTGAACTGGGAAAGACAAAGAAAGTTCCAGCAG 540
Db 1170 AAGCGCATTTGGAACCCGTACTACAACTGAACTGGGAAAGACAAAGAAAGTTCCAGCAG 1229
QY 541 AAACAGAGCCTTGAGACTTCAAGAGATCCGAGCGAGATTTGCCCAAGGGCCAGCGGTC 600
Db 1230 AAACAGAGCCTTGAGACTTCAAGAGATCCGAGCGAGATTTGCCCAAGGGCCAGCGGTC 1289
QY 601 GCGCCCTATMACACACGAGCTTCTCATGATGATCAGACAGGAGAGCGCGGATCTC 660
Db 1290 GCGCCCTATMACACACGAGCTTCTCATGATGATCAGACAGGAGAGCGCGGATCTC 1349
QY 661 AAAACCGGCTGTACTCCAAAGGGGGGGGGCCGCAATCCGAGACACCGAGCATGACGAC 720
Db 1350 AAAACCGGCTGTACTCCAAAGGGGGGGGGCCGCAATCCGAGACACCGAGCATGACGAC 1409
QY 721 TTCATGGAAGAGGGGGGTGAGAGGATGGGGGACGATGGGATGGAGGGAGCGGACGAC 780
Db 1410 TTCATGGAAGAGGGGGGTGAGAGGATGGGGGACGATGGGATGGAGGGAGCGGACGAC 1469
QY 781 GAGTTTCTGACGCGGACTTCTCGAGACGTACGAGCGGTACACACGAGAGCCTGCAG 840
Db 1470 GAGTTTCTGACGCGGACTTCTCGAGACGTACGAGCGGTACACACGAGAGCCTGCAG 1529
QY 841 AACATGAGCAGCAGAGGCTCATCAAGAGTACCTGGAACCTGGAGAACTGGCTTCGCGGC 900
Db 1530 AACATGAGCAGCAGAGGCTCATCAAGAGTACCTGGAACCTGGAGAACTGGCTTCGCGGC 1589
QY 901 ATGAGAGAGCAGAAACACCGGCTGCGGCTGAGAGACCAAGCGGCTGGTGGCAGCAGCGC 960
Db 1590 ATGAGAGAGCAGAAACACCGGCTGCGGCTGAGAGACCAAGCGGCTGGTGGCAGCAGCGC 1649
QY 961 CGTGTGCGGAGCTGAGCTGAGCTGAGCGGCTGCGCGCGGAGAACTCCAGCTGCTG 1020

Db 1550 CGTGTGGGGAGCTGGAGCTGGAGCTGGAGCGGCTGGCCGGAGAACCTTCAGCTGCTG 1709
QY 1021 ACCGAGAACGAAGTGCACCGGCGACGAGCGAGCGCGCTTTTCCAAATTGGAGACTAG 1080
Db 1710 ACCGAGAACGAAGTGCACCGGCGACGAGCGAGCGCGCTTTTCCAAATTGGAGACTAG 1769

RESULT 4

US-09-918-995-403/C
Sequence 403, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
PRIORITY FILING DATE: 2001-07-30
PRIORITY FILING DATE: 1999-01-20
PRIORITY FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 403
LENGTH: 495
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(495)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-403

Query Match 34.1%; Score 368.8; DB 9; Length 495;
Best Local Similarity 99.5%; Pred. No. 7.1e-89;
Matches 370; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 709 AGCATATGACGACTTCATGGAAGAGGGGTGAGAGAGATGGGGCAGCGATGGATGGGA 768
Db 495 AGCATATGACGACTTCATGGAAGAGGGGTGAGAGAGATGGGGCAGCGATGGATGGGA 436
QY 769 GGGAGCGGACGAGTTTCTGACGCGGACTTCTCGAGAGCTGAGAGCGGTACACACG 828
Db 435 GGGAGCGGACGAGTTTCTGACGCGGACTTCTCGAGAGCTGAGAGCGGTACACACG 376
QY 829 GAGAGCTGCGAGAACATGAGCAAGCAGAGCTCATCAAGAGTACTGGAACCTGGAGAAG 888
Db 375 GAGAGCTGCGAGAACATGAGCAAGCAGAGCTCATCAAGAGTACTGGAACCTGGAGAAG 316
QY 889 TGCCCTTCGCGCATGAGAGAGAGAGCAACCGGCTGCGGCTGAGAGAGCAACCGGCTGGT 948
Db 315 TGCCCTTCGCGCATGAGAGAGAGAGCAACCGGCTGCGGCTGAGAGAGCAACCGGCTGGT 256
QY 949 GGGAGCGGACGAGTTTCTGACGCGGACTTCTCGAGAGCTGAGAGCGGTGCGCGGAGAAC 1008
Db 255 GGGAGCGGACGAGTTTCTGACGCGGACTTCTCGAGAGCTGAGAGCGGTGCGCGGAGAAC 196
QY 1009 CTCGAGCTGCTGACCGAGAACGAACTGACCGGCGAGAGAGAGCGCGCTTTCCAAAG 1068
Db 195 CTCGAGCTGCTGACCGAGAACGAACTGACCGGCGAGAGAGAGCGCGCTTTCCAAAG 136
QY 1069 TTTGGAGACTAG 1080
Db 135 TTTGGAGACTAG 124

RESULT 5

US-09-983-965-4895
Sequence 4895, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengping
APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
PRIORITY FILING DATE: 2001-10-26
PRIORITY FILING DATE: 1999-12-15
PRIORITY FILING DATE: 1999-12-15
PRIORITY FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 4895
LENGTH: 414
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 24-LIB34-008-Q1-E1-F7
US-09-983-965-4895

Query Match 33.6%; Score 362.8; DB 10; Length 414;
Best Local Similarity 92.3%; Pred. No. 2.7e-87;
Matches 382; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 462 GAGACGCGCCCTCCAGAGAAAGCGGCAATGGAAACGCTACTACAGCTGAAGTGGAGAGA 521
Db 1 GAGACGCGCCCTCCAGAGAAAGCGGCTTGGAAACGCTACTACAGCTGAAGTGGAGAGA 60
QY 522 GAGAGAAAGTTGACGAGAGAAACAGAGCCTTTCAGAGTTCAGAGATCCGACCGAGATGT 581
Db 61 GAGAGAAAGTTGACGAGAGAAACAGAGCCTTTCAGAGTTCAGAGATCCGACCGAGATGT 120
QY 582 CGCCAGAGGCGAGCGGCTGCGCCCTTATACACACGAGTTCCTCATGTGATGACGA 641
Db 121 CGCCAGAGGCGAGCGGCTGCGCCCTTATACACACGAGTTCCTCATGTGATGACGA 180
QY 642 CCAGAGAGGCGGATCCAAACCGGCTGTACTCCAGAGGCGGCGCCGCAATGCGA 701
Db 181 CCAGAGAGGCGGATCCAAACCGGCTGTACTCCAGAGGCGGCGCCGCAATGCGA 240
QY 702 CGACACGAGGATGACGACTTCATGGAAGAGGGGTGAGAGAGATGGGGCAGCGATGG 761
Db 241 CGACACGAGGATGACGACTTCATGGAAGAGGGGTGAGAGAGATGGGGCAGCGATGG 300
QY 762 GATGAGAGGAGCGGACGAGCGAGTTTCTGACGCGGACTTCTCGAGACGTAACGCGGTA 821
Db 301 GATGAGAGGAGCGGACGAGCGAGTTTCTGACGCGGACTTCTCGAGACGTAACGCGGTA 360
QY 822 CCACAGGAGAGCGCTGCGAGAACATGAGCAAGCAGAGCTCATCAAGAGTACCT 875
Db 361 CCACAGGAGAGCGCTGCGAGAACATGAGCAAGCAGAGCTCATCAAGAGTACCT 414

RESULT 6

US-09-954-531-813
Sequence 813, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
FILE REFERENCE: 689290-77
CURRENT APPLICATION NUMBER: US/09/954,531
PRIORITY FILING DATE: 2002-05-02
PRIORITY FILING DATE: 2000-09-18
PRIORITY FILING DATE: 2000-09-18
PRIORITY FILING DATE: 2000-09-20
PRIORITY FILING DATE: 2000-09-20
PRIORITY FILING DATE: 2000-09-20
PRIORITY FILING DATE: 2000-09-20
PRIORITY FILING DATE: 2000-09-22
PRIORITY FILING DATE: 2000-09-22

```

: NUMBER OF SEQ ID NOS: 1392
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 813
: LENGTH: 461
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(461)
: OTHER INFORMATION: n-a,t,g or c
US-09-954-531-813

Query Match      32.9%; Score 355.2; DB 9; Length 461;
Best Local Similarity 95.7%; Pred. No. 3e-85;
Matches 374; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 1 ATGGCCAGGACCATCTTGTGACGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGT 60
   |||||
DB 72 ATGGCCAGGACCATCTTGTGACGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGT 131

QY 61 GCTGCTCTGTCCAGGAAGACTGAACCTTGAGCGCCCCAGCGCGGAGAGCGGGTG 120
   |||||
DB 132 GCTGCTCTGTCCAGGAAGACTGAACCTTGAGCGCCCCAGCGCGGAGAGCGGGTG 191

QY 121 CCCGAGGAGGACAGTAGTGGAATCGAGAGCGTTCCCGCAGTTGGGGGCGGTCGGGG 180
   |||||
DB 192 CCCGAGGAGGACAGTAGTGGAATCGAGAGCGTTCCCGCAGTTGGGGGCGGTCGGGG 251

QY 181 CCGGAGGAGGAAAGGAGCCTGGAATCCCAACACCTCCCTTCAGAGCCAGGCTGTCCA 240
   |||||
DB 252 CCGGAGGAGGAAAGGAGCCTGGAATCCCAACACCTCCCTTCAGAGCCAGGCTGTCCA 311

QY 241 GAATCTAGCTGCTGCTGAAGAGAGGCGAGAAAGGCGCCAGATGGGGACACTGTCGGCTGC 300
   |||||
DB 312 GAATCTAGCTGCTGCTGAAGAGAGGCGAGAAAGGCGCCAGATGGGGACACTGTCGGCTGC 371

QY 301 GGGGACTTCCCGCGCGCGGCGAGAAAGTGGAACCGAGCGCGCGAGCTGCTGCCCGAG 360
   |||||
DB 372 -GGGACTTCCCGCGCGCGGCGAGAAAGTGGAACCGAGCGCGCGAGCTGCTGCCCGAG 430

QY 361 CTTGTGATGACTCCGAGGCGAGTAGTGG 391
   |||||
DB 431 CTTGTGATGACTCCGAGGCGAGTAGTGG 461

RESULT 7
US-10-291-230-14/c
: Sequence 14, Application US/10291230
: Publication No. US20030108939A1
: GENERAL INFORMATION:
: APPLICANT: Ruffner, Duane E.
: APPLICANT: Pierce, Michael L.
: APPLICANT: Chen, Zhidong
: TITLE OF INVENTION: Directed Antisense Libraries
: FILE REFERENCE: T6678.US.A
: CURRENT APPLICATION NUMBER: US/10/291,230
: CURRENT FILING DATE: 2002-11-07
: PRIOR APPLICATION NUMBER: US 09/647,344
: PRIOR FILING DATE: 2000-12-04
: PRIOR APPLICATION NUMBER: PCT/US99/06742
: PRIOR FILING DATE: 1999-03-28
: PRIOR APPLICATION NUMBER: US 60/079,792
: PRIOR FILING DATE: 1998-03-28
: PRIOR APPLICATION NUMBER: US 60/107,504
: PRIOR FILING DATE: 1998-11-06
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 14
: LENGTH: 8705
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: pshuttle
```

```

US-10-291-230-14

Query Match      6.3%; Score 68.4; DB 9; Length 8705;
Best Local Similarity 44.0%; Pred. No. 1.6e-08;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGCGGAGAGGAGTGGGGACAGCAGCAGACAGCTGGGGAAGAAAAACCTAAGAGA 465
   |||
DB 7982 GGGGCGGAGAGGAGTGGGGACAGCAGCAGACAGCTGGGGAAGAAAAACCTAAGAGA 7923

QY 466 CGCCCTCTCAAGAAGACCGCATTTGAAACCTACTACAACTGAACTGGGAAGAG 525
   |||
DB 7922 GAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7863

QY 526 AAAAGTTGACAGGAAACAGAGCTTTCGAGCTTCAAGATCCGAGCCGAGATGTTGCC 585
   |||
DB 7862 CAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7803

QY 586 AAGGGCCAGGCGGTCGCGCCCTTAAACACACGAGTTCCATGATGATCAGACAG 645
   |||
DB 7802 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7743

QY 646 GAGGAGCCGAGATCTCAAAACCGGCTGTACTCCAAAGCGGCGCGCCGCAATCCGAGAC 705
   |||
DB 7742 GAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7683

QY 706 ACCAGCATGACACTTCAATGGAAGAGGGGTGAGAGATGGGGGACCGATGGCATG 765
   |||
DB 7682 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7623

QY 766 GGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 825
   |||
DB 7622 GAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7563

QY 826 ACGGAGAGCCTGCAAGAACATGAGCAACAGAGCTCATCAAGAGTACTGGAACCTGAG 885
   |||
DB 7562 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7503

QY 886 AAGTGCCTTCGCGCATGAGAGGAGCAGAACACCGGCTGCGGCTGAGAGAGCAGCGCTG 945
   |||
DB 7502 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7443

QY 946 GGTGCGCAGCAGCGCGGCTGTGCGGAGCTGAGAGCTGAGAGTGAACCGGCTGCGCGGAG 1005
   |||
DB 7442 GAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7383

QY 1006 AACCTCAGCTGTGACCCGAGAACGAACTGCACCAGGAGGAGGAGGAGGAGGAGGAG 1059
   |||
DB 7382 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7329

RESULT 8
US-09-954-531-1352
: Sequence 1352, Application US/09954531
: Patent No. US20020165180A1
: GENERAL INFORMATION:
: APPLICANT: Weaver, Zoe
: TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
: FILE REFERENCE: 689290-77
: CURRENT APPLICATION NUMBER: US/09/954,531
: CURRENT FILING DATE: 2002-05-02
: PRIOR APPLICATION NUMBER: US/60/233,133
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US/60/234,009
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,034
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,509
: PRIOR FILING DATE: 2000-09-22
: PRIOR APPLICATION NUMBER: US/60/234,567
: PRIOR FILING DATE: 2000-09-22
: NUMBER OF SEQ ID NOS: 1392
```

SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1352
; LENGTH: 4270
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-1352

Query Match 6.3%; Score 67.8; DB 9; Length 4270;
Best Local Similarity 49.9%; Pred. No. 2e-08;
Matches 171; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 712 GATGACGACTTATGGAAGAGGGGTGAGAGATGGGGGACGATGGATGGAGGG 771
DB 439 GAGTACGACTTCAAGATGCGACAGCGCTGAGAGAGCTCCGAGCAGCAGCAGCAA 498
QY 772 GACGCGACGAGCTTCTGAGCGGAGCTTCTGGAGACCTAGAGCGGATACCAACGAG 831
DB 499 GTGGCGCTCTACAGCTGAGCTGAGCGAGCTACCAAGGCTGAGAGCGCCCAAG 558
QY 832 AGCCGTGACAGACATGAGCAAGCAGAGCTCATCAAGAGTACCTGTAAGAGTGC 891
DB 559 CTGAGCTCTGACCAAGACAGAGCGGCGCTGCGAGAGCTGAAGAGGCGC 618
QY 892 CTCTGGCGATGAGAGCAGACCAACCGGCTGCGGCTGAGAGCAGACGCGCTGGTGC 951
DB 619 CGCATGCGGCTGAGCTCCCTCAGCTACCAAGCTCCGCGCTCAGAGAGCAGCGCCAGTGC 678
QY 952 GACGAGCGGCGTGTGCGGAGCTGAGCTGAGAGCTGAGAGCGGCTGCGCGAGAACCTC 1011
DB 679 GCTGAAGATCGCATTCGGAGAGCTGAGAGAGCGCCATGCGCGGAGCGGAGCAAGTCCGG 738
QY 1012 CAGCTGCTGACGAGAGCAAGCACTGCAACCGCGAGCAGAGCGAG 1054
DB 739 AAGATGCTGAGCGCCAGAGAGAGATGATGAGAGATGCGGG 781

RESULT 9

US-09-880-107-2427
; Sequence 2427, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2427
; LENGTH: 4270
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M94362
US-09-880-107-2427

Query Match 6.3%; Score 67.8; DB 10; Length 4270;
Best Local Similarity 49.9%; Pred. No. 2e-08;
Matches 171; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 712 GATGACGACTTATGGAAGAGGGGTGAGAGATGGGGGACGATGGATGGAGGG 771
DB 439 GAGTACGACTTCAAGATGCGACAGCGCTGAGAGAGCTCCGAGCAGCAGCAGCAA 498
QY 772 GACGCGACGAGCTTCTGAGCGGAGCTTCTGGAGACCTAGAGCGGATGACAGCGAG 831

DB 499 GTGGCGCTCTACAGCTGAGCTGAGCAGACCTACCAAGGCCAGCTGAGACGCCCAAG 558
QY 832 AGCCGTGACAGACATGAGCAAGCAGAGCTCATCAAGAGTACCTGTAAGAGTGC 891
DB 559 CTGAGCTCTGACCAAGACAGAGCGGCGCTGCGAGAGCTGAAGAGGCGC 618
QY 892 CTCTGGCGATGAGAGCAGACCAACCGGCTGCGGCTGAGAGCAGACCGCTGGTGC 951
DB 619 CGCATGCGGCTGAGCTCCCTCAGCTACCAAGCTCCGCGCTCAGAGAGCAGCGCCAGTGC 678
QY 952 GACGAGCGGCGTGTGCGGAGCTGAGCTGAGAGCTGAGACCGGCTGCGCGAGAACCTC 1011
DB 679 GCTGAAGATCGCATTCGGAGAGCTGAGAGAGCGCCATGCGCGGAGCGGAGCAAGTCCGG 738
QY 1012 CAGCTGCTGACGAGAGCAAGCACTGCAACCGCGAGCAGAGCGAG 1054
DB 739 AAGATGCTGAGCGCCAGAGAGAGATGAGAGATGCGGG 781

RESULT 10

US-09-962-832-225
; Sequence 225, Application US/09962832
; Patent No. US20020110821A1
; GENERAL INFORMATION:
; APPLICANT: Ederer, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-74
; CURRENT APPLICATION NUMBER: US/09/962,832
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,077
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,280
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 259
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 225
; LENGTH: 2108
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-832-225

Query Match 5.6%; Score 60.8; DB 10; Length 2108;
Best Local Similarity 51.1%; Pred. No. 1.3e-06;
Matches 143; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 789 GCAGCGGAGCTCTGCGAGAGCTAGAGCGGATGACCAACGAGAGCTGCAAGACATGAG 848
DB 782 GCAGCTGAGCTCTGAGACAGCAGAGGAGACCTGAAAGCAGCTGAGAGCAGCAGAGAGG 841
QY 849 CAGCAGAGAGCTCATCAAGAGAGTACCTGGAAGTGGAGAGTCCCTGCGCGATGAGAGA 908
DB 842 GCAGCTGAGAGTCCAGAGAGAGCAGATGAGGAGCTGAAGTACCTGGAAGCAGAGAGAGG 901
QY 909 CGAGAACCAACCGGCTGCGGCTGAGAGAGCAGAGCGGTGGGTGCGAGCAGAGCGCGTGC 968
DB 902 GCAGCTGAGAGCAGCTGATGATCAGCAGAGAGCAGCAGCAAGCTCCAGAGCAGCAGATGG 961
QY 969 GGAGCTGAGCTGAGAGCTGAGAGCGGCTGCGCGCGAGAACCTCCAGCTGAGACAGAGAA 1028
DB 962 GCAGCTGAGAGCAGCTGAGAGCAGCAGAGAGAGGAGCGCTTAAGCATCTGAGAGCAGAGAGG 1021
QY 1029 CGAAGTGCACCGGCGAGCAGAGAGCGAGCGCGCTTCCAG 1068
DB 1022 GCAAGTGGAGAGCTGAGAGAGAGCAGAGAGGCGAGCTGAAG 1061

RESULT 11

US-09-854-133-157
; Sequence 157, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.

APPLICANT: Mohamath, Raodoh
APPLICANT: Henderson, Robert A.
APPLICANT: Benson, Darin R.
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 210121.475C10
CURRENT APPLICATION NUMBER: US/09/854.133
CURRENT FILING DATE: 2001-05-11
NUMBER OF SEQ ID NOS: 735
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 157
LENGTH: 2313
TYPE: DNA
ORGANISM: Homo sapien
US-09-854-133-157

Query Match 5.4%; Score 58.4; DB 9; Length 2313;
Best Local Similarity 52.5%; Pred. No. 5.6e-06;
Matches 128; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 817 CGGTACCACACGAGAGCCCTGCAGAACATGAGACAGAGAGCTCATCAAGAGTACTG 876
DB 968 CGGCTGCAGGCGCGAGGGAAGACGCTCCCTACACAGAGCTGAAGCAGAAACGCGCGAG 1027
QY 877 GAAGTGGAGAAAGTCCCTTCGCGCATGAGAGAGAGAACAAACCGCTGCGGTGGAGAGC 936
DB 1028 CACGGGAGAGAGCGGAGCGCGCCCGCGCCCAAGAGAGAGAGTCTGCGGCTGCGAG 1087
QY 937 AAGCGGCTGGTGGCGACGAGCGCGCTGTGCGGAGCTGAGCTGAGACCGCTG 996
DB 1088 CAGCTGCAGAGAGAGAGAGCGGAGCGAGCTGACAGAGCTGGAGCTGCTGCGAGGCGAG 1147
QY 997 CGGCGGAGAACCTCCAGCTGCTGACCGAGAACAGACTGACCGGAGAGAGCGAGC 1056
DB 1148 CGGCAAGCGAGCGGCTGCTGCGAGAGAGAGAGAGAGCGCGCAGACCGCGAG 1207
QY 1057 CCGC 1060
DB 1208 CTGC 1211

RESULT 12
US-09-738-973-157
Sequence 157, Application US/09738973
Patent No. US20020110563A1
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Henderson, Robert A.
APPLICANT: Lodes, Michael J.
APPLICANT: Fling, Steven P.
APPLICANT: Mohamath, Raodoh
APPLICANT: Algate, Paul A.
APPLICANT: Secrist, Heather
APPLICANT: Indirias, Carol Joseph
APPLICANT: Benson, Darin R.
APPLICANT: Elliot, Mark
APPLICANT: Mannion, Jane
APPLICANT: Kalos, Michael D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 210121.475C9
CURRENT APPLICATION NUMBER: US/09/738.973
CURRENT FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 587
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 157
LENGTH: 2313
TYPE: DNA
ORGANISM: Homo sapien
US-09-738-973-157

Query Match 5.4%; Score 58.4; DB 10; Length 2313;

Best Local Similarity 52.5%; Pred. No. 5.6e-06;
Matches 128; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 817 CGGTACCACACGAGAGCCCTGCAGAACATGAGACAGAGAGCTCATCAAGAGTACTG 876
DB 968 CGGCTGCAGGCGCGAGGGAAGACGCTCCCTACACAGAGCTGAAGCAGAAACGCGCGAG 1027
QY 877 GAAGTGGAGAAAGTCCCTTCGCGCATGAGAGAGAGAACAAACCGCTGCGGTGGAGAGC 936
DB 1028 CACGGGAGAGAGCGGAGCGCGCCCGCGCCCAAGAGAGAGAGTCTGCGGCTGCGAG 1087
QY 937 AAGCGGCTGGTGGCGACGAGCGCGCTGTGCGGAGCTGAGCTGAGACCGCTG 996
DB 1088 CAGCTGCAGAGAGAGAGAGCGGAGCGAGCTGACAGAGCTGGAGCTGCTGCGAGGCGCGAG 1147
QY 997 CGGCGGAGAACCTCCAGCTGCTGACCGAGAACAGAACTGACCGGAGAGAGCGAGC 1056
DB 1148 CGGCAAGCGAGCGGCTGCTGCGAGAGAGAGAGAGCGCGCAGACCGCGAG 1207
QY 1057 CCGC 1060
DB 1208 CTGC 1211

RESULT 13
US-09-764-868-12
Sequence 12, Application US/09764868
Patent No. US2002016871A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT232
CURRENT APPLICATION NUMBER: US/09/764.868
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 1510
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 2314
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-868-12

Query Match 5.4%; Score 58.4; DB 9; Length 2314;
Best Local Similarity 52.5%; Pred. No. 5.6e-06;
Matches 128; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 817 CGGTACCACACGAGAGCCCTGCAGAACATGAGACAGAGAGCTCATCAAGAGTACTG 876
DB 976 CGGCTGCAGGCGCGAGGGAAGAGCTCCCTACACAGAGACTGAAGCAGAAACGCGCGAG 1035
QY 877 GAAGTGGAGAAAGTCCCTTCGCGCATGAGAGAGAGAACAAACCGCTGCGGTGGAGAGC 936
DB 1036 CAGCGGAGAGAGAGGAGCGCGCCCGCGCCCAAGAGAGAGAGAGCTGCGGTGCGAG 1095
QY 937 AAGCGGCTGGTGGCGACGAGCGCGCTGTGCGGAGAGCTGAGCTGAGACCGCTG 996
DB 1096 CAGCTGCAGAGAGAGAGAGAGAGAGCTGCGAGAGCTGAGAGCTGCTGCGAGAGCGCGAG 1155
QY 997 CGGCGGAGAACCTCCAGCTGCTGACCGAGAACAGAACTGACCGGAGAGAGCGAGC 1056
DB 1156 CGGCAAGCGAGCGGCTGCTGCGAGAGAGAGAGAGAGCGCGCAGCAGCAGCGAG 1215
QY 1057 CCGC 1060
DB 1216 CTGC 1219

RESULT 14
US-10-043-487-122
Sequence 122, Application US/10043487
Publication No. US20030055220A1
GENERAL INFORMATION:

APPLICANT: HYBRIGENICS
APPLICANT: PIERRE, LEGRAIN
TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptide
TITLE OF INVENTION: mammalian polypeptides
FILE REFERENCE: B4778A
CURRENT APPLICATION NUMBER: US/10/043,487
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/261,130
PRIOR FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 561
SOFTWARE: PatentIn version 3.1
SEQ ID NO 122
LENGTH: 607
TYPE: DNA
ORGANISM: Shigella Flexneri
US-10-043-487-122

Query Match 5.0%; Score 53.6; DB 9; Length 607;
Best Local Similarity 52.2%; Pred. No. 8.1e-05;
Matches 119; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 841 AACATGACGACGAGCTCATCAAGAGTACCTGGAAGTGAAGAGTCCCTCTGCGC 900
DB 230 AAGATGAGGACAGCATTAAGAGAGGCGGAACCTAGCGAGATTAAACCGTCGACATC 289
QY 901 ATGAGAGACGAGAACACCGGCTGCGCTGAGAGCAGCAGCGCTGGGTGGCGAGCGC 960
DB 290 ACTGAGTCACTCACTCCATCCGTCGCTGAGAGAGGCCAGAACGAGCGGATGACCGT 349
QY 961 CCGTGGCGGAGCTGAGCTGAGCTGAGCAGCGGCTGCGCGAGAACTCCAGCTGCTG 1020
DB 350 CTTGTGCTGGTGAGAGAGGCTCTGCAACAGCTGGCAGAAAGAGAGGCCCAAGCTGCTG 409
QY 1021 ACCGAGAACGAACTGCACGCGCAGCAGAGCGCGCTTTCCAAAG 1068
DB 410 GCCCAGCTGGCCAGAGTGTCAAGAGCAGCGGCGGAGGCTCCCCAG 457

RESULT 15
US-10-043-487-105
Sequence 105, Application US/10043487
Publication No. US20030055220A1
GENERAL INFORMATION:
APPLICANT: HYBRIGENICS
APPLICANT: PIERRE, LEGRAIN
TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptide
TITLE OF INVENTION: mammalian polypeptides
FILE REFERENCE: B4778A
CURRENT APPLICATION NUMBER: US/10/043,487
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/261,130
PRIOR FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 561
SOFTWARE: PatentIn version 3.1
SEQ ID NO 105
LENGTH: 1609
TYPE: DNA
ORGANISM: Shigella Flexneri
US-10-043-487-105

Query Match 5.0%; Score 53.6; DB 9; Length 1609;
Best Local Similarity 52.2%; Pred. No. 9.9e-05;
Matches 119; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 841 AACATGACGACGAGCTCATCAAGAGTACCTGGAAGTGAAGAGTCCCTCTGCGC 900
DB 1232 AAGATGAGGACAGCATTAAGAGAGGCGGAACCTAGCGAGATTAAACCGTCGACATC 1291
QY 901 ATGAGAGACGAGAACACCGGCTGCGCTGAGAGCAGCAGCGCTGGGTGGCGAGCGC 960
DB 1292 ACTGAGTCACTCACTCCATCCGTCGCTGAGAGAGGCCAGAACGAGCGGATGACCGT 1351
QY 961 CCGTGGCGGAGCTGAGCTGAGCTGAGCAGCGGCTGGCGGAGAACTCCAGCTGCTG 1020

DB 1352 CTTGTGCTGGGACGAGCAGGCTCTGCAACAGCTGGCAGAAAGAGAGCCCAAGCTGCTG 1411
QY 1021 ACCGAGAACGAACTGCACCGGAGCAGCAGGAGCGCGCTTTCCAAAG 1068
DB 1412 GCCCAGCTGGCCAGAGTGTCAAGAGCAGCGGCGGAGGCTCCCCAG 1459

Search completed: June 18, 2003, 05:28:17
Job time: 964 secs

[illegible]

PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX Montano M, Wiltman B;
XX WPI, 2002-519107/55.
DR P-PSDB: ABB76495.
XX
XX Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,
PT useful for the prevention, diagnosis and treatment of e.g. breast
PT cancer, testicular cancer, prostate cancer, uterine cancer, cervical
PT cancer, ovarian cancer and colon cancer -
XX
XX Claim 1; Fig 1A-B; 52p; English.
XX
XX The present sequence is the coding sequence for human oestrogen
CC downregulated gene 1 (EDG1), a tumour suppressor gene that is
CC downregulated by oestrogen in mammary epithelial cells. The gene
CC was identified by yeast two-hybrid screenings for oestrogen
CC receptor-interacting proteins in breast epithelial cells. It
CC was localised to chromosome arm 17q. EDG1 mRNA expression is
CC prevalent in normal mammary epithelial cells and in other human
CC hormone-responsive tissues such as the ovary, prostate and testis.
CC Expression is low in breast cancer epithelial cells. Oestradiol,
CC which induces breast cancer cell growth, has an inhibitory effect
CC on EDG1 mRNA expression in breast cancer cells. Hexamethylene
CC bis-acetamide, an inducer of differentiation and apoptosis,
CC upregulates EDG1 mRNA expression in breast cancer cells. The
CC invention provides EDG1 polynucleotides and polypeptides. In a
CC claimed method, a test sample from an individual suspected of
CC having, or known to have breast, testicular, prostate, uterine,
CC cervical, ovarian or colon cancer is assayed for EDG1 transcript
CC using a polynucleotide that is complementary to the present
CC sequence or by RT-PCR using a primer derived from the present
CC sequence. A decrease in the level of transcript compared to the
CC level in a test sample indicates that the test sample contains or
CC was derived from cancerous cells antibody. A claimed method for
CC decreasing the proliferation of breast, prostate, testicular,
CC ovarian, uterine, cervical or colon cancer cells involves increasing
CC EDG1 protein activity in the cells, either by contacting the cells
CC with EDG1 protein or its fragment or functional equivalent, or with
CC a nucleic acid encoding EDG1 protein, its fragment or functional
CC equivalent.
XX
XX Sequence 1080 BP; 265 A; 296 C; 376 G; 143 T; 0 other:
SQ
Query Match 100.0%; Score 1080; DB 24; Length 1080;
Best Local Similarity 100.0%; Pred. No. 8-7e-214;
Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 CCTTGTCATGACTCCGAGGAGAGTAAAGTTGGGGGCTCCCTGCCGACGGGGGAGAGAG 420
DB 361 CCTTGTCATGACTCCGAGGAGAGTAAAGTTGGGGGCTCCCTGCCGAGGGGGAGAGAGAG 420
QY 421 TGGGACAGCAGCAGAGCAGTGGGGGAAGAAAAACATTAAGACAGCCCTCCAAAG 480
DB 421 TGGGACAGCAGCAGAGCAGTGGGGGAAGAAAAACATTAAGAGACGCCCTCCAAAG 480
QY 481 AAGCGGCAATTGGAAACCGTACTACAAAGTGAACCTGGGAAGAGAAAAAGTTGACAG 540
DB 481 AAGCGGCAATTGGAAACCGTACTACAAAGTGAACCTGGGAAGAGAAAAAGTTGACAG 540
QY 541 AAACAGAGCCTTGAGCTTCAAGATCCGAGCCGAGATGTTCCCAAGGGCCAGCGGTC 600
DB 541 AAACAGAGCCTTGAGCTTCAAGATCCGAGCCGAGATGTTCCCAAGGGCCAGCGGTC 600
QY 601 GGGCCCTTAAACACACAGCAGTCTCATGATGATCAAGCCAGAGAGAGCCGATGTC 660
DB 601 GGGCCCTTAAACACACAGCAGTCTCATGATGATCAAGCCAGAGAGAGCCGATGTC 660
QY 661 AAAACCGGCTGTACTCCAAAGCGGCGCCGCAATCCGACGACACCCAGGATGACGAC 720
DB 661 AAAACCGGCTGTACTCCAAAGCGGCGCCGCAATCCGACGACACCCAGGATGACGAC 720
QY 721 TTCAATGAAAGAGGGGTGAGAGAGATGGGGCAGCATGGAGTGGAGGGAGCGCAC 780
DB 721 TTCAATGAAAGAGGGGTGAGAGAGATGGGGCAGCATGGAGTGGAGGGAGCGCAC 780
QY 781 GAGTTTGCAGCGGAGCTTTCGAGAGCTACGAGCGGTACACAGGAGAGCCTGAG 840
DB 781 GAGTTTGCAGCGGAGCTTTCGAGAGCTACGAGCGGTACACAGGAGAGCCTGAG 840
QY 841 AACATGACCAACAGCAGAGCTCATCAAGAGTACTGGAATCGAGAAAGTGCCTCGGC 900
DB 841 AACATGACCAACAGCAGAGCTCATCAAGAGTACTGGAATCGAGAAAGTGCCTCGGC 900
QY 901 ATGAGAGAGAGAAACACCGGCTGCGGCTGGAGAGCAAGCGGCTGGTGGAGACAGCG 960
DB 901 ATGAGAGAGAGAAACACCGGCTGCGGCTGGAGAGCAAGCGGCTGGTGGAGACAGCG 960
QY 961 CGTGGCGGAGCTGAGCTGAGCTGAGACCGGCTGCGGCGGCGAGAACCTCAGTGTG 1020
DB 961 CGTGGCGGAGCTGAGCTGAGCTGAGACCGGCTGCGGCGGCGAGAACCTCAGTGTG 1020
QY 1021 ACCGAGAAAGAACTGACACCGGAGAGAGAGCGCGCTTCCAAAGTTGGAGACTAG 1080
DB 1021 ACCGAGAAAGAACTGACACCGGAGAGAGAGCGCGCTTCCAAAGTTGGAGACTAG 1080

RESULT 2
AAV82778
ID AAV82778 standard; cDNA; 2199 BP.
XX
XX AAV82778;
AC
AC 25-FEB-1999 (first entry)
DT
DT
XX
XX
DE Clone bp783_3 isolated from human foetal kidney cDNA library.
XX
XX Secreted protein; nutritional activity; immune stimulating; vaccine;
OS suppressing activity; haematopoiesis regulating activity;
XX tissue growth activity; activin; inhibin activity; chemotaxis;
XX chemokine activity; haemostasis; thrombolytic activity; receptor;
XX ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
XX tumour inhibition; gene therapy; ds.
XX Homo sapiens.
OS
OS W09842739-A2.
PN
PN 01-OCT-1998.
PD
PD 20-MAR-1998; 98WO-US05653.
PF

XX 19-MAR-1998: 98US-0044466.
 PR 21-MAR-1997: 97US-0822167.
 XX (GEMV) GENETICS INST INC.
 XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Werberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX WPI: 1998-609890/51.
 DR P-PSDB: AAW85455.
 XX New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, placenta or adult
 PT pineal gland cDNA libraries.
 XX
 PS Claim 1: Page 66-67: 113pp: English.
 CC The present sequence encodes a secreted protein. The polynucleotide and
 CC secreted protein are predicted to have biological activities which would
 CC make them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data is given.
 CC Suggested activities include nutritional activity, immune stimulating
 CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity (no data is
 CC given in the specification to support these activities). The
 CC polynucleotide is also stated to be useful for gene therapy.
 XX
 SO Sequence 2199 BP: 552 A; 511 C; 674 G; 462 T; 0 other:
 Query Match 99.7%: Score 1076.8; DB 19; Length 2199;
 Best Local Similarity 99.8%: Pred. No. 4.5e-213;
 Matches 1078: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCGAGCATTCTGTGAGATATCAACACCGCTCAACTAGCACTGTACAGGT 60
 DB 707 ATGGCCGAGCATTCTGTGAGATATCAACACCGCTCAACTAGCACTGTACAGGT 766
 QY 61 GCTGCTGCTGTCAGAGAGAGTGAACCTTGAGCGCCCCAGCGCGGAGAGCGGGTG 120
 DB 767 GCTGCTGCTGTCAGAGAGAGTGAACCTTGAGCGCCCCAGCGCGGAGAGCGGGTG 826
 QY 121 CCGGAGGAGACAGTAGTGGCAATCGAGAGCGTTCCCGCAAGTTGGTGGCCCTCGGGG 180
 DB 827 CCGGAGGAGACAGTAGTGGCAATCGAGAGCGTTCCCGCAAGTTGGTGGCCCTCGGGG 886
 QY 181 CCGGAGGAGGAGAGGAGGCTGGAATCCCAACCCCTCCCTTGCAGACCCAGGCGTGTCCA 240
 DB 887 CCGGAGGAGGAGAGGAGGCTGGAATCCCAACCCCTCCCTTGCAGACCCAGGCGTGTCCA 946
 QY 241 GAATCTAGCTGCTGAGAGAGGCGGAGAAAGGCGCAGAAATGGGAGCAGCTGCTGGCTGGC 300
 DB 947 GAATCTAGCTGCTGAGAGAGGCGGAGAAAGGCGCAGAAATGGGAGCAGCTGCTGGCTGGC 1006
 QY 301 GGGGACTTCCCGCGCGCGCGGAGAGAGTGAACCGACCGCCGACAGCTGCTGCCAG 360
 DB 1007 GGGGACTTCCCGCGCGCGCGGAGAGAGTGAACCGACCGCCGACAGCTGCTGCCAG 1066
 QY 361 CTTTGTATGACTTCGAGGCGCAGTAAGTTGGGGGCTCTGCGGAGGGGCGAAGAGAGAG 420
 DB 1067 CTTTGTATGACTTCGAGGCGCAGTAAGTTGGGGGCTCTGCGGAGGGGCGAAGAGAGAG 1126
 QY 421 TGGGGAACAGCAGAGAGACAGCTGGGGAGAAAAAACATAGAGACGCCCTGTCCAAGAG 480
 DB 1127 TGGGGAACAGCAGAGAGACAGCTGGGGAGAAAAAACATAGAGAGCGCCGTGTCCAAGAG 1186
 QY 481 AAGCGGATTGGAAACCGTACTACAGCTGAAGTGGGAGAGAGAGAAAAAGTTTCGAGAG 540
 DB 1187 AAGCGGATTGGAAACCGTACTACAGCTGAAGTGGGAGAGAGAAAAAGTTTCGAGAG 1246

QY 541 AAACAGACCTTGAGCTTCAGAGATCCGAGCCGAGATGTTCCCAAGGCGCAGCGGTC 600
 DB 1247 AAACAGACCTTGAGCTTCAGAGATCCGAGCCGAGATGTTCCCAAGGCGCAGCGGTC 1306
 QY 601 GCGCCCTATTAACACACGACGAGTTCTTCATGATGATTCAGACAGCAGAGAGCGGATCTC 660
 DB 1307 GCGCCCTATTAACACACGAGCTTCATGATGATTCAGACAGCAGAGAGCGGATCTC 1366
 QY 661 AAAACCGGCTGTACTTCACAGCGGCGCGCCCAAAATCCGACAGACACACCATGATAGAGAC 720
 DB 1367 AAAACCGGCTGTACTTCACAGCGGCGCGCCCAAAATCCGACAGACACACCATGATAGAGAC 1426
 QY 721 TTGATGAAGAAGGCGGTGAGAGGATGGGCGGACGATGGATGGAGGGAGCGGACG 780
 DB 1427 TTGATGAAGAAGGCGGTGAGAGGATGGGCGGACGATGGATGGAGGGAGCGGACG 1486
 QY 781 GAGTTTCTGACGCGGACCTTCTGAGACCTAGCAGCGGTATCCACACGAGACCTTGACG 840
 DB 1487 GAGTTTCTGACGCGGACCTTCTGAGACCTAGCAGCGGTATCCACACGAGACCTTGACG 1546
 QY 841 AACATGAGCAAGCAGAGAGCTCATCAGAGATACCTGGAAGTGGAGAGTCCCTCGCGC 900
 DB 1547 AACATGAGCAAGCAGAGAGCTCATCAGAGATACCTGGAAGTGGAGAGTCCCTCGCGC 1606
 QY 901 ATGAGAGACGAGAAACACCGGCTGCGGCTGAGAGACGAGCGGCTGGGTGCGACAGACGCG 960
 DB 1607 ATGAGAGACGAGAAACACCGGCTGCGGCTGAGAGACGAGCGGCTGGGTGCGACAGACGCG 1666
 QY 961 CGTGTGCGGAGCTGAGAGCTGGAAGCTGGACCGGCTGCGCGCAGAGAACTCCAGCTGCTG 1020
 DB 1667 CGTGTGCGGAGCTGAGAGCTGGAAGCTGGACCGGCTGCGCGCAGAGAACTCCAGCTGCTG 1726
 QY 1021 ACCGAGACGAACTGCAACCGGCGCAGACGAGACGAGCGCGCTTTCCAAGTTTGGAGACTAG 1080
 DB 1727 ACCGAGACGAACTGCAACCGGCGCAGACGAGACGAGCGCGCTTTCCAAGTTTGGAGACTAG 1786

RESULT 3
 ABQ92015
 ID ABQ92015 standard; cDNA: 2199 BP.
 XX
 AC ABQ92015;
 XX
 DT 04-OCT-2002 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 12.
 XX
 KW Human: cytosolic; antirheumatic; antiarthritic; vulnery; analgesic;
 KW antinflammatory; antibacterial; immunosuppressive; antiParkinsonian;
 KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
 KW antidiacer; fungicide; antidiabetic; antiasthmatic; antiallergic;
 KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
 KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
 KW stem cell; growth factor; nervous system disease; neuropathy;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW osteoporosis; severe combined immunodeficiency; SCID; infection;
 KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN US2002065394-A1.
 XX
 PD 30-MAY-2002.
 XX
 PF 22-DEC-2000; 2000US-0745763.
 XX
 PR 18-MAR-1998; 98US-0040963.
 XX
 PA (JACO/) JACOBS K.
 PA (MCCO/) MCCOY J M.
 PA (LAVA/) LAVALLIE E R.
 PA (COLL/) COLLINS-RACIE L A.
 PA (EVAN/) EVANS C.

PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17952; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide which comprises a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2178 BP; 525 A; 513 C; 678 G; 462 T; 0 other:
Query Match 99.6%; Score 1075.2; DB 22; Length 2178;
Best Local Similarity 99.7%; Pred. No. 9,6e-213;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 1016 GCGGACTTCCCGCCCGCCGAGAAAGTGGACCGAGCCCGAGCCGAGCTGCTGCCAG 1075
QY CTTGTATGACTCCGAGGCCAGTAAGTTGGGGCTCTGCCGAGGGGGCGAAGAGAG 420
DB 1076 CTTGTATGACTCCGAGGCCAGTAAGTTGGGGCTCTGCCGAGGGGGCGAAGAGAG 1135
QY 421 TGGGACACGACAGACAGACAGCTGGGGAAAGAAAAACATTAAGACAGCCCTCCAAAG 480
DB 1136 TGGGACACGACAGACAGACAGCTGGGGAAAGAAAAACATTAAGAGAGCCCTCCAAAG 1195
QY 481 AAGCGCATTTGGAAACCTTACTACAACTGAACTGGGAAGAAAGAAAGTTGCACAG 540
DB 1196 AAGCGCATTTGGAAACCTTACTACAACTGAACTGGGAAGAAAGAAAGTTGCACAG 1255
QY 541 AAACGAGCCTTTCGAGCTTCAAGATCCGAGCCGAGATGTTGCCAAGGCCAGCGGCTC 600
DB 1256 AAACGAGCCTTTCGAGCTTCAAGATCCGAGCCGAGATGTTGCCAAGGCCAGCGGCTC 1315
QY 601 GCGCCCTTAAACACCGACAGCTTCTCATGATGATCAAGACAGACAGAGAGCGGATCTC 660
DB 1316 GCGCCCTTAAACACCGACAGCTTCTCATGATGATCAAGACAGACAGAGAGCGGATCTC 1375
QY 661 AAACCGGCTTGTACTCCAAAGCGGCGCCGCAATCCGACGACACGACGATGACGAC 720
DB 1376 AAACCGGCTTGTACTCCAAAGCGGCGCCGCAATCCGACGACACGACGATGACGAC 1435
QY 721 TTCAATGAAAGAGGGGTGAGAGATGGGGGACCGATGGGATGGAGGGAGCGGACG 780
DB 1436 TTCAATGAAAGAGGGGTGAGAGATGGGGGACCGATGGGATGGAGGGAGCGGACG 1495
QY 781 GAGTTTCGACGCGGACTTTCGAGAGATGACGAGCGGTACACAGAGAGCGCTGAC 840
DB 1496 GAGTTTCGACGCGGACTTTCGAGAGATGACGAGCGGTACACAGAGAGCGCTGAC 1555
QY 841 AACATGACCAAGCAGAGCTTATCAAGAGATGACGAGTGAATCGAGAACTGCTTCGCGC 900
DB 1556 AACATGACCAAGCAGAGCTTATCAAGAGATGACGAGTGAATCGAGAACTGCTTCGCGC 1615
QY 901 ATGAGAGACGAGAACACCGGCTGCGGTGAGAGACCAAGCGGCTGGTGGAGACAGCG 960
DB 1616 ATGAGAGACGAGAACACCGGCTGCGGTGAGAGACCAAGCGGCTGGTGGAGACAGCG 1675
QY 961 CGTGGCGGAGCTGAGCTGAGCTGAGACGCGGCTGCGCGCGAGAACCTCAGCTGCTG 1020
DB 1676 CGTGGCGGAGCTGAGCTGAGCTGAGACGCGGCTGCGCGCGAGAACCTCAGCTGCTG 1735
QY 1021 ACCGAGAACGAACCTGACCGGACGAGAGCGAGCGCGCTTTCGAAGTTGGAGACTAG 1080
DB 1736 ACCGAGAACGAACCTGACCGGACGAGAGCGAGCGCGCTTTCGAAGTTGGAGACTAG 1795
RESULT 5
ABL90067/c
ID ABL90067 standard; cDNA, 1083 BP.
XX
AC ABL90067;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 629.
XX
DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX
XX antitumor; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX
XX cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX
XX neurological disease; infection; human; secreted protein; gene; ss.
OS Homo sapiens.
XX
XX W0200190304-A2.
XX
XX 29-NOV-2001.
XX

PF 18-MAY-2001; 2001MO-US16450.
XX
XX 19-MAY-2000; 2000US-205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
XX WPI: 2002-122018/16.
DR P-PSDB: ABB89658.
XX
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders -
XX
XX
XX Claim 4; SEQ ID NO 629; 2081bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB89040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.
XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcr_sequences.
XX
XX Sequence 1083 BP; 301 A; 290 C; 199 G; 286 T; 7 other;
SQ
Query Match 40.2%; Score 434.2; DB 24; Length 1083;
Best Local Similarity 98.9%; Pred. No. 1.7e-80;
Matches 444; Conservative 3; Mismatches 1; Indels 1; Gaps 1;
OY 632 ATGATCAGCAGCAGAGAGAGCGGATCTCAAAACCGGCTTACTCCAAAGCGGCCCG 691
DB 1083 ATGATCAGCAGCAGAGAGAGCGGATCTCAAAACCGGCTTACTCCAAAGCGGCCCG 1024
OY 692 CCAATCGAGACACCGGATGACGCTTCATGGAAGAAGGGGGTATGAGATGGGG 751
DB 1023 -CAATCGAGACACCGGATGACGCTTCATGGAAGAAGGGGGTATGAGATGGGG 965
OY 752 GCACGATGATGGAGAGAGAGCGGATGATGATGATGATGATGATGATGATGATGAT 811
DB 964 GCACGATGATGGAGAGAGAGCGGATGATGATGATGATGATGATGATGATGATGAT 905
OY 812 ACGAGCGGTACCAACGAGAGAGCGGATGATGATGATGATGATGATGATGATGATGAT 871
DB 904 ACGAGCGGTACCAACGAGAGAGCGGATGATGATGATGATGATGATGATGATGATGAT 845
OY 872 ACGGGAATGAGAGAGAGAGCGGATGATGATGATGATGATGATGATGATGATGAT 931
DB 844 ACGTGAAGTGAAGAGAGAGCGGATGATGATGATGATGATGATGATGATGATGAT 785
OY 932 AGAGCAAGCGGCTGGTGGAGAGAGCGGATGATGATGATGATGATGATGATGATGATGAT 991
DB 784 AGAGCAAGCGGCTGGTGGAGAGAGCGGATGATGATGATGATGATGATGATGATGATGAT 725
OY 992 GGTGGCGGCGAGAACTCTCAGCTGTGACCGAGAAAGAACTCACCAGGAGAGAGAGC 1051
DB 724 GGTGGCGGCGAGAACTCTCAGCTGTGACCGAGAAAGAACTCACCAGGAGAGAGAGC 665
OY 1052 GAGGCGCGCTTCCAGATTGGAGACTAG 1080
XXXXXXXXXXXXXXXXXXXXXXXXXXXX

DB 664 GAGCGCGCTTCCAGATTGGAGACTAG 636
RESULT 6
AAV86036
ID AAV86036 standard; CDNA: 523 BP.
XX
XX AAV86036;
AC
XX 27-APR-1999 (first entry)
DT
XX
DE EST clone B227.
XX
XX Expressed sequence tag; secreted protein; haematopoiesis regulator; tissue growth; activin; inhibin; tumour invasion suppressor; EST; human; chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytics; receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
XX Homo sapiens.
OS
XX WO9845435-A2.
FN
XX 15-OCT-1998.
PD
XX 10-APR-1998; 98WO-US06954.
PF
XX 10-APR-1997; 97US-0835913.
PR
XX (GENV) GENETICS INST INC.
PA
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
XX WPI: 1999-070076/06.
DR
XX
XX New polynucleotides encoding human secreted proteins - derived from e.g. human blood, kidney, foetal lung, placenta, testes, brain, ovary, pituitary, retina and colon cDNA libraries
XX
XX
XX Claim 1; Page 97; 633bp; English.
XX
XX This sequence represents an expressed sequence tag (EST), and is a CC polynucleotide of the invention. The polynucleotides of the invention are CC all secreted EST sequences isolated from a variety of human tissue CC sources. The EST sequences and proteins encoded by them are predicted to CC have useful biological activities which would make them suitable for CC treating, preventing or ameliorating medical conditions in humans and CC animals, although no supporting data is given. Suggested activities CC include nutritional activity, immune stimulating or suppressing activity, CC haematopoiesis regulating activity, tissue growth activity, CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition CC activity. The EST sequences are also stated to be useful for gene CC therapy.
XX
XX Sequence 523 BP; 112 A; 149 C; 153 G; 108 T; 1 other;
SQ
Query Match 36.6%; Score 395.8; DB 20; Length 523;
Best Local Similarity 96.6%; Pred. No. 1.3e-72;
Matches 403; Conservative 1; Mismatches 13; Indels 0; Gaps 0;
OY 1 ATGGCCGAGGCAATCTTGTGAGATATCAACACGAGCTCAAACTAGCAACTGACAGGT 60
DB 104 ATGGCCGAGGCAATCTTGTGAGATATCAACACGAGCTCAAACTAGCAACTGACAGGT 163
OY 61 GCTGCTGCTGTCACGAGAGAGCTAACCTGAGCGCCGCCAGCGCGGAGAGAGCGGCTG 120
DB 164 GCTGCTGCTGTCACGAGAGAGCTAACCTGAGCGCCGCCAGCGCGGAGAGAGCGGCTG 223
OY 121 CCGAGAGAGAGCTAGTAGTGCAATCGAGAGAGCTTCCCGAGTGGGTTGGCGGCGCGG 180
DB 224 CCGAGAGAGAGCTAGTAGTGCAATCGAGAGAGCTTCCCGAGTGGGTTGGCGGCGGCGG 283
XXXXXXXXXXXXXXXXXXXXXXXXXXXX

OY 181 CCGAGGGGGAAGGAGCCTGGAAATCCCAACCACTCCCTTGACAGCCAGGCGCTGTCCA 240
Db 284 CCGGAGGGGGAAGGAGCCTGGAAATCCCAACCACTCCCTTGACAGCCAGGCGCTGTCCA 343
OY 241 GAATCTAGCTGCTGCTGAGAGAGGGCGAAGAGGGCCGAATGGGAGACTGTCTCGCTGSC 300
Db 344 GAATCTAGCTGCTGCTGAGAGAGGGCGAAGAGGGCCGAATGGGAGACTGTCTCGCTGSC 403
OY 301 GGGCACTTCCCGCCCGGCGACAAAGTGGAAACCGAGCGCCGAGCGGAGCTGCTCGCCAG 360
Db 404 GGGCACTTCCCGCCCGGCGACAAAGTGGAAACCGAGCGCCGAGCGGAGCTGCTCGCCAG 463
OY 361 CCTTGTCATGACTCCGAGGCGCAGTAAGTTGGGGGCTCCTCGCCGAGGGGCGCAAG 417
Db 464 CCTTGTCATGACTCCGAGGCGCAGTAAGTGGGGTGCAGCGGCTCTTTGGCCCTCGAG 520
RESULT 7
ABL63543
ID ABL63543 standard; DNA; 461 BP.
XX ABL63543;
AC
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SFO ID NO:1880.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; kiltumour; cancerous;
KW cytoslastic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 22-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.

PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237423P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 1880; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (1) of a signature gene set, where (1)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL6164
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (1) has cytoslastic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 461 BP; 105 A; 134 C; 138 G; 80 T; 4 other;
OY Query Match 32.9%; Score 355.2; DB 24; Length 461;
OY Best Local Similarity 95.7%; Pred. No. 3e-64; Mismatches 16; Indels 1;
OY Matches 374; Conservative 0; Mismatches 16; Indels 1;
Db 1 ATGCCGAGCCATTCTTGTCAAGATATCAACACCAAGCCCAACTAGCAACTGTACAGGT 60
Db 72 ATGCCGAGCCATTCTTGTCAAGATATCAACACCAAGCCCAACTAGCAACTGTACAGGT 131
OY 61 GCTGCTGCTGTCCAGAGAGCTGAACCTTGAGCCGCCGCCGAGGAGGCGGGTG 120
Db 132 GCTGCTGCTGTCCAGAGAGCTGAACCTTGAGCCGCCGCCGAGGAGGCGGGTG 191
OY 121 CCCGAGAGAGAGAGTAGTGTGCAATCGAGAGGCTTCCCAATGGTGGTCCGCGGG 180
Db 192 CCCGAGAGAGAGAGTAGTGTGCAATCGAGAGGCTTCCCAATGGTGGTCCGCGGG 251
OY 181 CCGGAGGGGGAAGGAGCCTGGAATCCCAACCACTCCCTTGCAACCCAGGCTGTGTCA 240
Db 252 CCGGAGGGGGAAGGAGCCTGGAATCCCAACCACTCCCTTGCAACCCAGGCTGTGTCA 311
OY 241 GAATCTAGCTGCTGCTGAGAGAGGGCGAAGAGGGCCGAATGGGAGCACTGTCCCTGGC 300
Db 312 GAATCTAGCTGCTGCTGAGAGAGGGCGAAGAGGGCCGAATGGGAGCACTGTCCCTGGC 371
OY 301 GGGCACTTCCCGCCCGGCGACAAAGTGGAAACCGAGCGCCGAGCGGAGCTGCTGCCAG 360
Db 372 -GGCACTTCCCGCCCGGCGACAAAGTGGAGCCGAGCGCGGAGCGGAGCTGCTGSCAG 430
OY 361 CCTTGTCATGACTCCGAGGCGCAGTAAGTTGG 391

Db 431 CCTGTGATGACTCCGAGGCGAGTAGTGG 461
 |||
 RESULT 8
 AAV87316
 ID AAV87316 standard; cDNA: 342 BP.
 XX
 AC AAV87316;
 XX
 DT 27-APR-1999 (first entry)
 XX
 DE EST clone BP783.
 XX
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 OS Homo sapiens.
 XX
 PN MO9845435-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06954.
 XX
 PR 10-APR-1997; 97US-0835913.
 XX
 PA (GEMV) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX
 DR WPI: 1999-070076/06.
 XX
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 XX
 PS Claim 1; Page 526; 633pp; English.
 XX
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cachectin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.
 XX
 SQ Sequence 342 BP; 90 A; 79 C; 83 G; 89 T; 1 other;
 Query Match 12.8%; Score 138; DB 20; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.9e-19;
 Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCCGAGCCATTCTGTGATATCAACACAGCCTCAAACTAGCACTGTACAGT 60
 DB 196 ATGGCCGAGCCATTCTGTGATATCAACACAGCCTCAAACTAGCACTGTACAGT 255
 QY 61 GCTCTCTGTCTCCAGAGAGCTCAACCTTGAAGCCCCCGAGCGGAGAGCGGGTG 120
 DB 256 GCTCTCTGTCTCCAGAGAGCTCAACCTTGAAGCCCCCGAGCGGAGAGCGGGTG 315
 QY 121 CCCGAGGAGACACTAGG 138
 |||

Db 316 CCCGAGGAGACACTAGG 333
 |||
 RESULT 9
 AAH99667/C
 ID AAH99667 standard; cDNA: 997 BP.
 XX
 AC AAH99667;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human protein encoding cDNA sequence SEQ ID NO:502.
 XX
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiac; central nervous system; vitruclide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antanaemic; anaemia;
 KW antiaggregant; haemostatic; vulnery; antitumor; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nocotropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153455-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 22-DEC-2000; 2000WO-US35017.
 XX
 PR 23-DEC-1999; 99US-0471275.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 XX
 PR 25-APR-2000; 2000US-0552317.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-457603/49.
 XX
 PS P-PSDB; AAM25726.
 XX
 CC Isolated human polynucleotides encoding polypeptides, useful for the
 CC treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 CC Claim 1; Page 566-567; 1217pp; English.
 XX
 AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; vitruclide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antanaemic; antiaggregant; haemostatic; vulnery;
 CC antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nocotropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and

CC neurological disorders.

XX Sequence 997 BP; 184 A; 323 C; 266 G; 224 T; 0 other;

SQ Query Match 10.4%; Score 112.4; DB 22; Length 997;

Best Local Similarity 62.7%; Pred. No. 4.5e-14;

Matches 192; Conservative 0; Mismatches 11; Indels 3; Gaps 1;

QY 733 GGGGGTGGAGAGATGGGGCACCAGTGGATGGAGGGAGCGACCGAGTTTCGAC 792

DB 887 GGGGAAGTGTAGCCCGGACAGTGTATGGGGCGGACGCGACGCTGAGTTCACCGG 828

QY 793 CGGAGCTTCGAGAGCTGACGAGCGGTACACACGAGAGCCTGCGAATGAGCAG 852

DB 827 AAGGACTTCTCTAGACTTACGAACGCTTCCACACCGAGAGCCTGCGAGGCCCGCAGCAG 768

QY 853 CAGGAGCTCATCAAGAGTACCTGGAAGTGGAGAGTCTCTCCGCGCATGAGAGAG 912

DB 767 CAGGAGCTGTGTGCGAGACTACCTGAGCTGGAGAGAGCGGCTGTGCGAGCGGAGAGAG 708

QY 913 AACACCGGCTGGCGCTGGAGAGCAACCGGCTGGTGGCA---CGACGGCGGTGTGCGG 969

DB 707 ACTAGAGGCTGTGAGAGCTGACAGCGGTGCACCGCGCCAGCATGTCCTGCGCGAGGTGAG 648

QY 970 GAGCTGAGCTGGAGCTGAGACCGCGCTGCGCGAGCAACCTCCAGCTGTGACCGAGAAC 1029

DB 647 GAGCTGGCTGCGCAGAGTCCAGAGGCTCCGAGCCGAAACCGAGCGGCTTCCTCAGAGAAC 588

QY 1030 GAACGTG 1035

DB 587 CAGATG 582

RESULT 10

AAH12885/c

ID AAH12885 standard; cDNA; 567 BP.

XX AAH12885;

XX 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO:9720.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX BP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

XX Claim 3; SEQ ID 9720; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises at least 15 nucleotides, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

CC XX

SQ Sequence 567 BP; 182 A; 139 C; 87 G; 149 T; 10 other;

Query Match 7.7%; Score 82.8; DB 22; Length 567;

Best Local Similarity 76.7%; Pred. No. 5.2e-08;

Matches 122; Conservative 0; Mismatches 34; Indels 3; Gaps 2;

QY 922 CTGCGGCTGGAGAGCAGCGGCTGTGGCGACACCGCGGTGCGGAGCTGAGACTG 981

DB 539 CNGNTTCGTTGAGACANCGTTGGGTGGAAACGCGNCTTTGGCGGAG-TGAGATTG 481

QY 982 GAGCTGAGCGGCTGCGCGCGCGAGACCTTCAGCTGCTGACCGAAGCACTGCACCG 1041

DB 480 GAGTTG--ACCGGTGGCGCGCGGAGAACTTCAGTTGATGACGAGAAAGATTGCACCG 423

QY 1042 CAGCAGGAGGAGCGCGCGCTTCCAAAGTTGAGAGACTAG 1080

DB 422 CAGCAGGAGGAGCGCGCGCTTCCAAAGTTGAGAGACTAG 384

RESULT 11

AAV55831

ID AAV55831 standard; DNA; 799 BP.

XX AAV55831;

XX 18-NOV-1998 (first entry)

XX Nucleotide sequence of the stabilizing sequence-encoding insert.

XX Fusion protein; stabilizing polypeptide; proteolytic degradation;

XX resistance; half-life; autoimmune disease; inflammation; nitro drug;

XX IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;

XX nitroreductase protein; enzyme therapy; produg therapy; protease;

XX cancer; pathological condition; ss.

XX Epstein-Barr virus.

XX WO9822577-A1.

XX 28-MAY-1998.

XX 17-NOV-1997; 97WO-IB01508.

XX 25-JUN-1997; 97US-0048945.

XX 15-NOV-1996; 96US-0030986.

XX (MASU/) MASUCCI M G.

XX Masucci MG;

XX

DR WPI: 1998-312463/27.
 XX New fusion proteins resistant to proteolytic degradation -
 PT comprising a core protein with a stabilising polypeptide comprising
 PT a peptide sequence containing glycine repeats
 XX
 PS Disclosure: Fig 4B; 120pp; English.

CC This is a nucleotide sequence of the stabilising sequence-encoding
 CC insert. The invention provides a method for increasing the resistance
 CC of a core protein to proteolytic degradation that comprises linking or
 CC inserting onto or into the core protein a stabilising polypeptide of
 CC formula $(\text{Gly})_n(\text{Gly})_m(\text{Gly})_k$ where Glya, Glyb, Glyc are 1-6
 CC sequential Gly residues and x, y, z are Ala, Ser, Val, Ile, Leu, Met,
 CC Phe, Pro or Thr and n can be anything between 1-66. x, y and z need not
 CC be identical from n repeat to n repeat. Alternatively a nucleic acid
 CC encoding the stabilising polypeptide can be linked onto or inserted into
 CC a nucleic acid encoding a core protein. The fusion proteins of the
 CC invention are more resistant to degradation by proteases and, thus, have
 CC a longer half-life than the unfused core protein. The products can be
 CC used for treating autoimmune diseases, cancer and inflammation. In
 CC particular, the core protein may be an Ikappab regulator protein for the
 CC treatment of inflammatory bowel disease, or a nitroreductase protein
 CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
 CC or other pathological conditions. The fusion proteins can also be used in
 CC diagnostic methods such as in vivo imaging.

XX Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;

Query Match 6.3%; Score 68.4; DB 19; Length 799;

Best Local Similarity 44.0%; Pred. No. 5.2e-05; Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGCGAGAGAGTGGGAGACGACAGACAGTGGGGAAGAAAACTAAGAGA 465
 DB 63 GGGCAGAGACAGAGAGAGGGGCGACAGAGAGAGGGGCGAGAGAGGGGCGAG 122
 QY 466 CGCCCGTCCAGAAAGAGCGGATTGGAAACCGTACTACAGCTGAAGTGGAGAGAG 525
 DB 123 GAGGGGCGAGAGACAGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGGGGCGAG 182
 QY 526 AAAAAGTTGAGAGAGAAAGAGACCTTCAGCTCAAGATCCGACCCAGATGTTCCGC 585
 DB 183 CAGGAGCAGAGAGAGGGGCGAGAGAGAGAGGGGCGAGAGAGAGAGAGAGAG 242
 QY 586 AAGGGCCAGCGGTCGCGCTATATACACGACGATTCCTCATGATATACAGACGAG 645
 DB 243 GGGCAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
 QY 645 GAGGAGCCGAGATCTAAACCGGCTGTACTCCAGCGGGCGCGCCAAATCCAGCAG 705
 DB 303 GAGGGGCGAGAGAGAGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAG 362
 QY 705 ACCAGCGATGAGACTTATGAGAAAGGGGGTGAAGAGATGGGGGCGAGCGATGGGATG 765
 DB 363 CAGGAGCAGAGAGGGGCGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAG 422
 QY 765 GAGGAGGAGCGAGCAGATTCTGACGCGGACTTCTCGAGACGTTACGAGCGGTACCA 825
 DB 423 GAGGGGCGAGAGGGGCGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAG 482
 QY 826 ACAGGAGCCCTGACAGAACTGAGCAAGCAGAGAGCTCATAGAGAGTACCTGGAACTGAG 885
 DB 483 CAGGAGGGGCGAGAGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 542
 QY 886 AAGTGGCTCTCGCGGCTGAGAGAGAGCAACCGGCTCGGCTGAGAGAGCAAGCGGCTG 945
 DB 543 GAGCAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
 QY 946 GGTGGGAGCAGCGCGTGTGCGGAGCTGAGAGCTGAGACCTGAGACCGGCTGCGCGCGAG 1005
 DB 603 GAGGGGCGAGAGGGGCGAGAGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAG 662

QY 1006 AACCTCCAGCTGTCAGACGAGAACGAACTGCACCGGACAGAGAGCGCGG 1059
 DB 663 GGGCAGAGACAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAG 716

RESULT 12
 ID AAA50254
 AAA50254 standard; DNA; 1926 BP.

XX AAA50254;

XX 07-NOV-2000 (first entry)

XX Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.

XX EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;

XX gene therapy; ds.

XX Epstein-Barr virus.

XX WO20004778-A1.

XX 17-AUG-2000.

XX 11-FEB-2000; 2000WO-US03547.

XX 11-FEB-1999; 99US-0249585.

XX (PHAR-) PHARMACORPEIA INC.

XX Horlick RA, Chelsky D;

XX WPI: 2000-515062/46.

XX P-PSDB; AAY95856.

XX Stably transfecting eukaryotic cells with at least one episome for the

XX production of a desired protein in vitro and for gene therapy -

XX Disclosure: Fig 2; 53pp; English.

CC The present sequence is that of DNA encoding the Epstein-Barr virus
 CC (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is
 CC utilised in a novel method for obtaining a eukaryotic cell that is
 CC stably transfected with at least one episome. This method involves
 CC transfecting a eukaryotic cell with: (1) a first episome comprising
 CC an EBV origin of replication (oriP, see AAA50253), a gene encoding a
 CC first protein whose expression results in cell death and a
 CC selectable marker for eukaryotic cells; and (2) a second episome
 CC comprising an EBV oriP and a gene encoding a second protein, where
 CC expression of the second protein prohibits the occurrence of cell
 CC death resulting from expression of the first protein to produce
 CC doubly transfected cells which also express an antigen that
 CC promotes retention of the episomes by the cells. The doubly
 CC transfected cells are maintained under conditions in which the
 CC first and second proteins and the selectable marker are expressed,
 CC and the selective pressure specified by the marker is maintained.
 CC Under these conditions, only cells containing both episomes live.
 CC Preferably, EBNA1 is expressed from 1 of the episomes, and the
 CC protein of interest from the other episome. Either or both
 CC epitopes may further comprise a nucleic acid sequence encoding a
 CC protein desired to be expressed in the cell (e.g. a therapeutic
 CC protein), a nucleic acid encoding an RNA that is not intended to
 CC be translated (e.g. a therapeutic RNA), or a DNA sequence used as
 CC a tag for the cells. The method is applicable to cell culture or
 CC intact organisms, for gene therapy. It allows the rapid
 CC establishment of eukaryotic cells that stably and reliably express
 CC a gene of interest, using a novel method of selection, and
 CC maintenance of that selection without the need for exogenous
 CC selection factors, such as antibiotics.

XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match 6.3%; Score 68.4; DB 21; Length 1926;

Best Local Similarity 44.0%; Pred. No. 6e-05;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
QY 406 GGGGGGGAAGAGAGTGGGACAGCAGACAGCTGGGGAAGAAAACTAAGACA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 GGGCAGAGCAGAGAGGGGCGAGAGCAGAGAGGGGCGAGGGGCGAG 364
QY 466 CGCCCTCCAAAGAACGCCATTGAAAACCTACTACAGCTGAACCTGGGAAGAA 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GAGGGCAGAGCAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGGGG 424
QY 526 AAAAAGTTGACGAGAAGAACAGACCTTCGAGCTTCAAGATCCAGCCGATGTCGCC 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 CAGAGCAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGGGGCGAGAG 484
QY 586 AAGGGCCAGCCGCTCCGCCCTTAAACACCAGCAGTTCCTCATGATGATCAGCAG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 GGGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGAGGG 544
QY 646 GAGGAGCCGATCTCAAAACCGGCTGTACTCCAGCGGGCGCCGCAAAATCCAGCAG 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 GAGGGCAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 604
QY 706 ACCAGCGATGACGACTTCAATGGAAGAAGGGGCTGAGAGAGATGGGGCAGCGATG 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 CAGGACAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGG 664
QY 766 GAGAGGGAGGGCGAGCGCACTTTCTGCAAGCCGAGCTTCCGAGACGTAACGCGTAC 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 GAGGGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 724
QY 826 ACAGAGAGCTTCAGAAACATGACAGCAGAGCAGCTCATCAAGATGAACTGAG 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 CAGGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 784
QY 886 AAGTGCCTCTCGCGATGAGAGCAGAAACACCGGCTGGGCTGAGAGCAACCGGCTG 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 GAGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGG 844
QY 946 GGTGGCAGAGAGCGCGCTGTGCGGAGCTGTGAGACTGGAGTCCGCGCTGCGCGGAG 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 845 GAGGGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 904
QY 1006 AACCTCCAGCTGTGACCGGAAGCAACTGCAACCGCAGCAGAGAGGAGCGCG 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 GGGCAGAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGAGAGGGGCGAGAGCAG 958
```

RESULT 13
AAF82902
ID AAF82902 standard: DNA; 1926 BP.
XX AAF82902:

29-JUN-2001 (first entry)

EBV tethering protein EBNA1 encoding DNA.

XX Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
XX Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
XX EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.
XX Epstein-Barr virus.

OS

FT Key Location/Qualifiers

FT CDS 1..1926 /*tag= a

PN MO200125484-A2.

PD 12-APR-2001.

PF 29-SEP-2000; 2000MO-US26908.

XX 01-OCT-1999; 99US-0410399.

XX (UNMT) UNIV MICHIGAN.

XX Robertson ES; Cotter MA;

XX WPI: 2001-281736/29.

XX P-PSDB; AAB62332.

XX A composition for use in gene therapy comprises an expression vector
XX PT that includes a nucleic acid sequence encoding a nucleic acid binding
XX protein

XX Disclosure; Fig 9C; 60pp; English.

XX The invention provides a composition comprising nucleic acid, histone H1
XX protein and expression vector operationally encoding a protein suitable
XX for tethering the nucleic acid to the histone H1 protein, where the
XX tethering protein is LANA. The composition is useful in aiding the
XX retention of the viral DNA in the host cell. The viral vector encodes a
XX protein suitable for tethering DNA to histone H1. Methods for screening
XX for compounds which are agonistic or antagonistic for the tethering of
XX viral proteins to histone H1 and DNA binding sites are useful for
XX developing the method of viral transfer. The composition has applications
XX to gene therapy, including the treatment of multiple sclerosis,
XX Parkinson's disease, Huntington disease and diabetes. The present
XX sequence represents the nucleotide sequence of the Epstein-Barr virus
XX (EBV) tethering protein EBNA1.

XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

XX Query Match 6.3%; Score 68.4; DB 22; Length 1926;

XX Best Local Similarity 44.0%; Pred. No. 6e-05;

XX Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
QY 406 GGGGGGGAAGAGAGTGGGACAGCAGACAGCTGGGGAAGAAAACTAAGACA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 GGGCAGAGCAGAGAGAGGGGCGAGAGCAGAGAGGGGCGAGGGGCGAG 364
QY 466 CGCCCTCCAAAGAACGCCATTGAAAACCTACTACAGCTGAACCTGGGAAGAA 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GAGGGCAGAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 424
QY 526 AAAAAGTTGACGAGAAGAACAGACCTTCGAGCTTCAAGATCCAGCCGATGTCGCC 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 CAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 484
QY 586 AAGGGCCAGCCGCTCCGCCCTTAAACACCAGCAGTTCCTCATGATGATCAGCAG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 GGGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGG 544
QY 485 GAGGAGCCGATCTCAAAACCGGCTGTACTCCAGCGGGCGCCGCAAAATCCAGCAG 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 GAGGGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 604
QY 545 GAGGGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 GAGGGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 604
QY 706 ACAGAGAGCTTCAGAAACATGACAGCAGAGCAGCTCATCAAGATGAACTGAG 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 CAGGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 664
QY 766 GAGAGGGAGGGCGAGCGCACTTTCTGCAAGCCGAGCTTCCGAGACGTAACGCGTAC 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 GAGGGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 724
QY 725 CAGGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 CAGGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 784
QY 826 ACAGAGAGCTTCAGAAACATGACAGCAGAGCAGCTCATCAAGATGAACTGAG 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 ACAGAGAGCTTCAGAAACATGACAGCAGAGCAGCTCATCAAGATGAACTGAG 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 886 AAGTGCCTCTCGCGATGAGAGCAGAAACACCGGCTGGGCTGAGAGCAACCGGCTG 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 AAGTGCCTCTCGCGATGAGAGCAGAAACACCGGCTGGGCTGAGAGCAACCGGCTG 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 946 GGTGGCAGAGAGCGCGCTGTGCGGAGCTGTGAGACTGGAGTCCGCGCTGCGCGGAG 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


CC protein-protein interactions between known proteins and for isolating,
CC cloning or characterizing unknown proteins. The method is also useful
CC for screening candidate bioactive agents that modulate the
CC protein-protein interaction between a bait protein and a test protein
CC and thus are useful for the identification of novel therapeutic drugs.
CC The present method employs mammalian cloning systems that are highly
CC stable, and designed to reduce the background signals frequently found
CC in other systems such as the yeast two-hybrid system. The current
CC sequence represents the coding sequence for the Epstein-Barr virus
CC nuclear antigen, the viral replication protein, which, on binding to the
CC Epstein-Barr origin of replication (see AI164274) on the bait vector,
CC allows vector replication to occur.
XX

SO Sequence 2580 BP; 632 A; 512 C; 1055 G; 381 T; 0 other;

Query Match 6.38; Score 68.4; DB 24; Length 2580;

Best Local Similarity 44.08; Pred. No. 6.3e-05;

Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
QY 406 GGGGGCGAAGAGAGTGGGGACAGCAGCAGACGACTGGGGAGAAAAACATAGAGA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 688 GGGCAGAGCAGCGAGGAGGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 CGCCCGTCCAAAGAAAGCGCGCATTTGAAACCGTACTACAAGCTGAACCTGGGAAGAGA 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 748 GAGGGGCGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 526 AAAAAGTTCCAGCAGAAACAGAGCCTTCGAGCTTCAAGATCCGAGCCGAGATTTGCC 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 808 CAGAGCAGAGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 586 AAGGGCCAGCGGTCGGCCCTATACACCACGAGTTCTCATGATGATCAGCAGCAG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 868 GGGCAGAGAGGGGCGAGGAGGAGGAGCAGAGCAGAGAGGAGGAGGAGGAGGAGGAG 927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 646 GAGGAGCCGAGTCTCAAAACCGCGCTTACTTCCAAGCGGGCCGCCCAATCCGAGCAG 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 928 GAGGGCCAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 706 ACCAGCGATGACGACTTCATGGAAGAAGGGGTTAGAGAGATGGGGCAGCGATGGCATG 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 988 CAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1047
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 766 GAGAGGAGCGCAGCGAGTTTCTGACGCGGACTTTCGGAGACTAGAGCGGTATACAC 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1048 GAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 826 ACGGAGAGCTGCAAGACATGACAGCAGAGACCTCATCAAGAGATCCTGGAACCTGAG 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 CAGGAGGAGCAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 886 AAGTGCCTTCGCGCATGAGAGCAGAGAACACCGGCTGCGGCTGAGAGCAGACCGGCTG 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 GAGCAGAGAGGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 946 GGTGGCAGCAGCGCGCTGTGCGGGAGCTGAGCTGAGCTGACCGGCTGCGCGCGCAG 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 GAGGGGCGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1006 AACCTCAGCTGTGACCGAGAAACGAACCTCAACCGAGCAGAGCAGCGCCG 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 GGGCAGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: June 18, 2003, 01:10:55
Job time : 321 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 21:37:42 ; Search time 1820 Seconds
(without alignments)
9610.509 Million cell updates/sec

Title: US-09-972-758a-1
Perfect score: 1080
Sequence: 1 atgagcagcattctgtc.....tttccaagtttgagactag 1080

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	68.1	876	9	AL552998
2	727.8	67.4	903	14	BQ221641
3	707.2	65.5	982	12	BF984049
4	694.6	64.3	771	13	BI871190
5	652.8	60.4	852	13	BI091005
6	639.2	59.2	797	12	BG819109

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES
AL552998	AL552998	AL552998 LTI_NFL006.PL2 Homo sapiens CDNA clone CS0D1072YL06 5	AL552998	AL552998	GI:12892417	EST.	human.	1 (bases 1 to 876)	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.	Full-length cDNA libraries and normalization	Unpublished (2001)	Contact: Genoscope	Genoscope - Centre National de Sequencage
								BP 191 91006 EVRY cedex - France					Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
								Location/Qualifiers					1. 876
								/organism="Homo sapiens"					/db_xref="taxon:9606"
								/clone="CS0D1072YL06"					/clone_lib="LTI_NFL006.PL2"
								/tissue_type="Placenta"					/note="Vector: PCWVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end

ALIGNMENTS

C	7	628	58.1	1001	9	AL578242	AL578242
8	615.2	57.0	57.0	871	12	BG291150	BG291150
9	598	55.4	750	10	AV726891	AV726891	AV726891
10	585	54.2	949	12	BG165450	BG165450	BG165450
11	580.6	53.8	1294	13	BM474968	BM474968	BM474968
12	579	53.6	986	9	AL576112	AL576112	AL576112
13	572.6	53.0	1080	13	BM479496	BM479496	BM479496
14	552.2	51.1	557	13	BM129736	BM129736	BM129736
15	539.2	49.9	745	10	BE547281	BE547281	BE547281
16	525.2	48.6	933	12	BF674587	BF674587	BF674587
17	509.8	47.2	843	12	BF026064	BF026064	BF026064
18	498.2	46.1	907	12	BF181535	BF181535	BF181535
19	494.2	45.8	810	12	BF607249	BF607249	BF607249
20	488.4	45.2	563	10	AM665096	AM665096	AM665096
21	487.2	45.1	994	9	AL555845	AL555845	AL555845
22	486.4	45.0	762	12	BF167480	BF167480	BF167480
23	483.8	44.8	999	13	BI411874	BI411874	BI411874
24	477.4	44.2	482	9	AI948609	AI948609	AI948609
25	476.6	44.1	566	10	AM962738	AM962738	AM962738
26	471.2	43.6	547	9	AI796944	AI796944	AI796944
27	466.4	43.2	666	12	BG684490	BG684490	BG684490
28	438.4	40.6	907	13	BI412550	BI412550	BI412550
29	436.4	40.4	442	9	AI948617	AI948617	AI948617
30	432	40.0	638	9	AA546538	AA546538	AA546538
31	402	37.2	442	12	BF082480	BF082480	BF082480
32	391.6	36.3	462	12	AV728223	AV728223	AV728223
33	375.4	34.8	485	9	AI811680	AI811680	AI811680
34	372.2	34.5	450	10	AM166154	AM166154	AM166154
35	370.2	34.3	459	12	BF446806	BF446806	BF446806
36	368.6	34.1	577	13	BM440009	BM440009	BM440009
37	365.6	33.9	819	14	BQ222807	BQ222807	BQ222807
38	363.6	33.7	488	14	W34238	W34238	W34238
39	360.6	33.4	540	10	AM143910	AM143910	AM143910
40	358	32.9	366	10	AV708410	AV708410	AV708410
41	355.2	32.1	461	14	N21081	N21081	N21081
42	342.6	31.7	980	13	BI526654	BI526654	BI526654
43	331	30.6	734	12	BG122671	BG122671	BG122671
44	330	30.6	400	9	AL362139	AL362139	AL362139
45	330	30.6	480	9	AA928789	AA928789	AA928789

QY 631 GATGATCAGACAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 690
 Db 481 GATGATCAGACAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 540
 QY 691 GCGAATCGGACAGACAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 750
 Db 541 GCGAATCGGACAGACAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 600
 QY 751 GCGAATCGGACAGACAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 810
 Db 601 GCGAATCGGACAGACAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 660
 QY 811 TACGAGCGGTACACAGAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 869
 Db 661 TACGAGCGGTACACAGAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 720
 QY 870 GTACCTGGAGTGGAGAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 921
 Db 721 GTACCTGGAGTGGAGAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGGCGGCGCC 775

RESULT 3
 BP984049 982 bp mRNA linear EST 23-JAN-2001
 LOCUS 602306989F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4398238 5',
 DEFINITION mRNA sequence.
 ACCESSION BP984049
 VERSION BP984049.1 GI:12386861
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH_MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10100 row: a column: 23
 High quality sequence stop: 691.
 Location/Qualifiers
 1..982
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4398238"
 /clone_lib="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-Sport6;
 Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
 oligo-dt primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC library."
 BASE COUNT 261 a 235 c 320 g 166 t
 ORIGIN

Query Match 65.5% Score 707.2 DB 12 Length 982:
 Best Local Similarity 99.4% Pred No. 3.3e-134;
 Matches 720; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 357 CCAACCTTTGATGATCTCGAGGCGCACTAGTTGGGGCTCTGCGCCAGAGGGCGCAAGA 416
 Db 7 CCAACCTTTGATGATCTCGAGGCGCACTAGTTGGGGCTCTGCGCCAGAGGGCGCAAGA 66
 QY 417 GGAGTGGGAGACAGACAGACAGCTGGGGAAGAAAAACATAGAGAGAGCGCCGTCAA 476

Db 67 GGAGTGGGAGACAGACAGACAGCTGGGGAAGAAAAACATAGAGAGAGCGCCGTCAA 126
 QY 477 GAAGAGCGGCTTTGGAAACCGTACTACAAAGCTGAACCTGGGAAGAGAAAAAGTTGCA 536
 Db 127 GAAGAGCGGCTTTGGAAACCGTACTACAAAGCTGAACCTGGGAAGAGAAAAAGTTGCA 186
 QY 537 CGAAGAAACAGAGCGCTTGAGAGCTCAAGGATTCGAGCGGAGATGTTGCCAAGGCCAGCC 596
 Db 187 CGAAGAAACAGAGCGCTTGAGAGCTCAAGGATTCGAGCGGAGATGTTGCCAAGGCCAGCC 246
 QY 597 GGTGCGGCCCTTATACACACAGCAGTTCCTCATGATGATCAAGACACAGAGAGAGCCGA 656
 Db 247 GGTGCGGCCCTTATACACACAGCAGTTCCTCATGATGATCAAGACACAGAGAGAGCCGA 306
 QY 657 TCTCAAAACCGGCTGTACTCCAAAGCGGGGCGCCGCAATTCGACAGACACAGCGATGA 716
 Db 307 TCTCAAAACCGGCTGTACTCCAAAGCGGGGCGCCGCAATTCGACAGACACAGCGATGA 366
 QY 717 CGACTTCATGAGAAAGAGGGGTGAGAGAGATGGGGCAGCGATGGATGGGAGGGAGCG 776
 Db 367 CGACTTCATGAGAAAGAGGGGTGAGAGAGATGGGGCAGCGATGGATGGGAGGGAGCG 426
 QY 777 CAGCGAGTTTCTGACAGCGGAGCTTCTCGAGACGTAGCAGCGGTAACACAGAGAGAGCT 836
 Db 427 CAGCGAGTTTCTGACAGCGGAGCTTCTCGAGACGTAGCAGCGGTAACACAGAGAGAGCT 486
 QY 837 GCAGAACATGAGCAAGCAGAGAGCTTCATCAAGAGAGCTGGAATCGAGAGAGTCCCTCTC 896
 Db 487 GCAGAACATGAGCAAGCAGAGAGCTTCATCAAGAGAGCTGGAATCGAGAGAGTCCCTCTC 546
 QY 897 GCGCATGAGAGACAGAAACAACCGGCTGCGCTGAGAGAGCAAGCGGTGGGTGGCGAGA 956
 Db 547 GCGCATGAGAGACAGAAACAACCGGCTGCGCTGAGAGAGCAAGCGGTGGGTGGCGAGA 606
 QY 957 CGCGGTGTGCGGGAGCTGGAGCTGGAGCTGGAGCGGCTGCGCGCGAGAACTCCAGCT 1016
 Db 607 CGCGGTGTGCGGGAGCTGGAGCTGGAGCTGGAGCGGCTGCGCGCGAGAACTCCAGCT 666
 QY 1017 GCTACCGAGAACAGACTGACCGCGAGAGAGAGCGGCTGCTTCCAAAGTTGGAGA 1076
 Db 667 GCTACCGAGAACAGACTGACCGCGAGAGAGAGCGGCTGCTTCCAAAGTTGGAGA 726
 QY 1077 CTAG 1080
 Db 726 CTAG 729

RESULT 4
 B1871190 771 bp mRNA linear EST 11-OCT-2001
 LOCUS 603395081F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5404724 5',
 DEFINITION mRNA sequence.
 ACCESSION B1871190
 VERSION B1871190.1 GI:16044865
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH_MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Db 365 GGGGCGCGATGATGGAGGGAGCGGAGGAG-TTCTGCGAGGGGAGCTTCTCGAGGA 423
 Oy 809 CGTACGAGCGGTACACACGAGAGGCTTCAGACATGAGCAAGAGAGCTCATCAAG 868
 Db 424 CGTACGAGCGGTACACACGAGAGGCTTCAGACATGAGCAAGAGAGCTCATCAAG 483
 Oy 869 AGTACCTGGAAGAGAGTCTCTGCGGATGAGAGAGAGCAACCGGCTGCGGC 928
 Db 484 AGTACCTGGAAGAGAGTCTCTGCGGATGAGAGAGAGCAACCGGCTGCGGC 543
 Oy 929 TGGAGAGCAAGCGGTGAGGAGAGAGCGGCTGCGGAGAGTGAAGTGAAGTGG 988
 Db 544 TGGAGAGCAAGCGGTGAGGAGAGAGCGGCTGCGGAGAGTGAAGTGAAGTGG 603
 Oy 989 ACCGGTGGCGCGGAGAGACCTCCAGCTGTGACCCAGAACGAACTGCAACCGGACAGG 1048
 Db 604 ACCGGTGGCGCGGAGAGACCTCCAGCTGTGACCCAGAACGAACTGCAACCGGACAGG 662
 Oy 1049 AGCGAGCGCGCTTCCAGGTTTGAAGACTAG 1080
 Db 663 AGCGAGCGCGCTTCCAGGTTTGAAGACTAG 694

RESULT 6 797 bp mRNA linear EST 22-MAY-2001
 BG819109
 LOCUS 602781202P1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4932025
 DEFINITION 5', mRNA sequence.

ACCESSION BG819109
 VERSION BG819109.1 GI:14166696
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 797)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: LLAM10857 row: C column: 02
 High quality sequence stop: 797.

FEATURES

source
 1. 797
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4932025"
 /clone_1ib="NCI CGAP_Brn67"
 /tissue_type="anaplastic Oligodendroglioma with 1p/19q
 loss"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Brain; Vector: PCMV-SPOK6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.3 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 BASE COUNT 207 a 223 c 248 g 119 t
 ORIGIN

Query Match 59.2%; Score 639.2; DB 12; Length 797;
 Best Local Similarity 96.5%; Pred. No. 2,4e-120;
 Matches 708; Conservative 0; Mismatches 18; Indels 8; Gaps 5;
 Oy 1 ATGCCGAGCCATTCTTGTGATATATACACACAGGCTCAACTGACACTGACAGT 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 69 ATGCCGAGCCATTCTTGTGATATATACACACAGGCTCAACTGACACTGACAGT 128
 Oy 61 GCTGCTGCTTCCAGAGAGAGCTGAACCTGAACGCCCCCAGAGCCGAGAGAGCGGTTG 120
 Db 129 GCTGCTGCTTCCAGAGAGAGCTGAACCTGAACGCCCCCAGAGCCGAGAGAGCGGTTG 188
 Oy 121 CCCAGAGAGAGAGCTGAGAGAGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 180
 Db 189 CCCAGAGAGAGAGCTGAGAGAGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 248
 Oy 181 CCCAGAGAGAGAGCTGAGAGAGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 240
 Db 249 CCCAGAGAGAGAGCTGAGAGAGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 308
 Oy 241 GAATCTAGCTCTGAG 300
 Db 309 GAATCTAGCTCTGAG 368
 Oy 301 GAGGACTTCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 Db 369 GAGGACTTCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
 Oy 361 CCTGTGATGACTCCGAG 420
 Db 429 CCTGTGATGACTCCGAG 488
 Oy 421 TGGGGACAGCAGCAG 480
 Db 489 TGGGGACAGCAGCAG 548
 Oy 481 AAGCGGATTTGAAACCGTACTTCAAGCTGAACTGGGAAAGAGAGAGAGAGAGAGAG 538
 Db 549 AAGCGGATTTGAAACCGTACTTCAAGCTGAACTGGGAAAGAGAGAGAGAGAGAGAG 607
 Oy 539 AGAAGAGAGAGCTTCAAGAGCTTCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 598
 Db 608 AGAAGAGAGAGCTTCAAGAGCTTCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 667
 Oy 599 TCGCGGCTATTAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
 Db 668 TCGCGGCTATTAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 727
 Oy 658 CTCAAAAAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
 Db 728 CTCAAAAAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 783
 Oy 718 GACTTCATGAGAGA 731
 Db 784 GACTTCATGAGAGA 797

RESULT 7 1001 bp mRNA linear EST 16-FEB-2001
 AL578242/C
 LOCUS AL578242 LTI_NFL006.PL2 Homo sapiens cDNA clone CSDDK002YF07 3
 DEFINITION prime, mRNA sequence.

ACCESSION AL578242.1 GI:12942132
 VERSION AL578242.1
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1001)
 AUTHORS L.L.W.B., Gruber C., Jesse J., and Polayes D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1. 1001
 Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002X707"
/clone_1ib="LTL_NFL006_PL2"
/tissue_type="placenta"
/Note="Vector: PCWSPORT 6; Site:1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCWSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact: Feng Liang Life Technologies,
Rockville, Maryland 20850, USA Fax: (1) 301 610 8371
Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com"
BASE COUNT      218 a      289 c      271 t      5 others
ORIGIN

```

```

Query Match      58.1%; Score 628; DB 9; Length 1001;
Best Local Similarity 98.8%; Pred. No. 4.6e-118;
Matches 639; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

434 AGAGACAGCTGGGAGAGAAAAAATAGAGAGCCGCTCCAAAGAAAGCGGATTGGA 493
1001 AGAGACAGCTGGGAGAGAAAAAATAGAGAGCCGCTCCAAAGAAAGCGGATTGGA 942
494 AACCTACTACAGCTGAGCTGGAGAGAGAAAAAATTGAGAGAAACAGAGCTTC 553
941 AACCTACTACAGCTGAGCTGGAGAGAGAAAAAATTGAGAGAAACAGAGCTTC 882
554 GAGCTTCAAGAGTCCGAGCCGAGATGTTGGCAAGGGCCAGCGGTCGCTTATACA 613
881 GAGCTTCAAGATTCGAGCCGAGATGTTGGCCAAAGGGCCAGCGGTCGCTTATACA 822
614 CCAGCAGTTCCTATGATGATGATCAGACAGAGAGAGAGCCGATCTCAAAACCGGCTG 673
821 CCAGCAGTTCCTATGATGATGATCAGACAGAGAGAGAGCCGATCTCAAAACCGGCTG 762
674 ACTTCAAGCGGGCGCGCCCAAAATCCGAGACACACCGAGTGCAGCTTCATGGAAGA 733
761 ACTTCAAGCGGGCGCGCCCAAAATCCGAGACACACCGAGTGCAGCTTCATGGAAGA 702
734 GGGGTGAGAGATGGGGGAGCGAGTGGATGGAGAGGGGAGCGACGAGTTCTGACAC 793
701 GGGGTGAGAGATGGGGGAGCGAGTGGATGGAGAGGGGAGCGACGAGTTCTGACAC 642
794 GGGACTTCTCGAGAGATGAGCGGTACACACGAGAGCTGCAAGACATGAGCAAC 853
641 GGGACTTCTCGAGAGATGAGCGGTACACACGAGAGCTGCAAGACATGAGCAAC 582
854 AGGAGCTCATCAAGAGATGAGCTGAGAACTGAGAAAGCTCTCGGGCATGGAGAGAGA 913
581 AGGAGCTCATCAAGAGATGAGCTGAGAACTGAGAAAGCTCTCGGGCATGGAGAGAGA 522
914 ACAACGGGCTGGCGCTGGAGAGACAGCGGCTGGGTGGCGAGACCGCGGTGTGGGGAGC 973
521 ACAACGGGCTGGCGCTGGAGAGACAGCGGCTGGGTGGCGAGCTACCGCGGTGTGGGGAGC 462
974 TGGAGCTGAGAGCTGAGACCGGCTGGCGGCGGAGAACCTCAGTGTGACCGAGAAACGAA 1033
461 TGGAGCTGAGAGCTGAGACCGCTGGCGGCGGAGAACCTCAGTGTGACCGAGAAACGAA 403
913 TGCACCGGAGAGAGAGAGCGCGCTTTCCAAAGTTTGGAGACTAG 1080
402 TGCACCGGAGAGAGAGAGCGCGCTTTCCAAAGTTTGGAGACTAG 356

```

```

RESULT 8
LOCUS      BG291150
DEFINITION 6038718271 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4515812 5',
ACCESSION  BG291150

```

```

VERSION      BG291150.1 GI:13048811
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE        NIH-MGC http://img.ncbi.nlm.nih.gov/
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLMI0406 row: d column: 21
High quality sequence stop: 730.
Location/Qualifiers
1..871

```

```

FEATURES
source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4515812"
/clone_1ib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: bladder; Vector: PCW-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT      210 a      230 c      324 g      107 t
ORIGIN

```

```

Query Match      57.0%; Score 615.2; DB 12; Length 871;
Best Local Similarity 91.9%; Pred. No. 1.9e-115;
Matches 750; Conservative 0; Mismatches 53; Indels 13; Gaps 9;

176 CGGGGCGGAGGGGAGAGAGAGCTGGAATCCAAACCACTCCCTTGACAGACCAAGCCCT 235
9 CGGGGCGGAGGGGAGAGAGAGCTGGAATCCAAACCACTCCCTTGACAGACCAAGCCCT 68
236 GTCCAGATATCTAGTGTGCTGAGAGAGGCGGAGAGGCGCAAGATGGGAGACTGCTCCG 295
69 GTCCAGATATCTAGTGTGCTGAGAGAGGCGGAGAGGCGCAAGATGGGAGACTGCTCCG 128
296 CTGGCGGCGAATCCCGCCCGCGGAGAAATGGAACCGAGCGCGGAGGCGGAGCTGCTCG 355
129 CTGGCGGCGAATCCCGCCCGCGGAGAAATGGAACCGAGCGCGGAGGCGGAGCTGCTCG 188
356 CCCAGCTTCTCATGACTCCGAGAGCGAGTAAGTTGGGGCTCTCTGCGCGAGGGGCGAAG 415
189 CCCAGCTTCTCATGACTCCGAGAGCGAGTAAG-TGGGGGCTCTCTGCGCGAGGGGCGAAG 247
416 AGGAGTGGGAGAGAGAGAGAGAGAGCTGGGGAGAGAAAAAATAAGAGAGCCCGCTCCA 475
248 AGGAGTGGGAGAGAGAGAGAGAGAGCTGGGGAGAGAAAAAATAAGAGAGCCCGCTCCA 307
476 AGAAGAGCGGCAATTGGAAACCGTACTACAACTGAAGTGGGAGAGAGAAAAAGTTGCG 535
308 AGAAGAGCGGCA-TGGAAACCGTACTACAACTGAAGTGGGAGAGAGAAAAAGTTGCG 366
536 ACAGAAACAGAGAGCTTGGAGCTTCAAGATCCGAGCGGAGATGTTGCCCAAGGGCGAGC 595
367 ACAGAAACAGAGAGCTTGGAGCTTCAAGATCCGAGCGGAGATGTTGCCCAAGGGCGAGC 426
596 CGGTGCGGCTTATAACACACAGAGCTTCTCATGATGATCAGACAGAGAGAGAGCGGG 655
427 CGGTGCGGCTTATAACACACAGAGCTTCTCATGATGATCAGACAGAGAGAGAGCGGG 486

```

Query Match	55.4%	Score 598	DB 10	Length 750
Best Local Similarity	88.0%	Prod. No. 5	8e-112	
Matches 680	Conservative	0	Mismatches 39	Indels 54
				Gaps 1
QY	300	CGCGGACCTTCCCGCGCGCGAGTGTGACACCGAGCGCGGAGCCGACAGTGTGGCCCA		359
Db	1	CGCGGACCTTCCCGCGCGCGAGTGTGAGCCGACCGCGGAGGCCGAGTGTCTCCGCCA		60

DNA library created by: *in-house*
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
 Plate: LLAM0250 row: k column: 07
 High quality sequence stop: 622.
 Location/Qualifiers

FEATURES

source

1. 949
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4456062"
 /clone_lib="NIH_MGC_89"
 /issue_type="Hydrenephroma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally; oligo-dt primed.
 Average insert size 1.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 243 a 257 c 335 g 114 t
 ORIGIN

Query Match 54.2%; Score 585; DB 12; Length 949;
 Best Local Similarity 90.8%; Pred. No. 2.7e-109;
 Matches 747; Conservative 0; Mismatches 60; Indels 16; Gaps 11;

228 CCAGGCGCTGTCAGATCTAGTCTGCTAGAGAGGCGGCAAGAGGCGCAATGGGAGCA 287
 2 CCAGGCGCTGTCAGATCTAGTCTGCTAGAGAGGCGGCAAGAGGCGCAATGGGAGCA 61
 288 CTCGTCGCTGCGGCGGCACTTCCCGCCGCGCAGAAAGTGGACCGACCGCGGCGCA 347
 62 CTCGTCGCTGCGGCGGCACTTCCCGCCGCGCAGAAAGTGGACCGACCGCGGCGCA 120
 348 GCTGTCGCGGCGGCACTTCCCGCCGCGCAGAAAGTGGACCGACCGCGGCGCA 407
 121 GCTGTCGCGGCGGCACTTCCCGCCGCGCAGAAAGTGGACCGACCGCGGCGCA 179
 408 GGGCGAAGAGAGTGGGAGCAGCAGCAGACAGCAGTGGGGAAGAAACATAGAGACG 467
 180 GGGCGAAGAGAGTGGGAGCAGCAGCAGACAGCAGTGGGGAAGAAACATAGAGACG 239
 468 CCGCTCCAAAGAAAGCGGCACTTGGAAACCGTACTACAGCTGAATGGGAGAGAGAA 527
 240 CCGCTCCAAAGAAAGCGGCACTTGGAAACCGTACTACAGCTGAATGGGAGAGAGAA 238
 528 AAAGTTCAGAGAGAAAGAGACCTTCGAGCTTCAGAGTCCGAGCCGAGATGTTCCCAA 587
 299 AAAGTTCAGAGAGAAAGAGACCTTCGAGCTTCAGAGTCCGAGCCGAGATGTTCCCAA 358
 588 GGGCGACCGGCTGCGGCGGCACTTGGAAACCGTACTACAGCTGAATGGGAGAGAGAA 647
 359 GGGCGACCGGCTGCGGCGGCACTTGGAAACCGTACTACAGCTGAATGGGAGAGAGAA 418
 648 GAGCGCGATCTCAAAACCGGCTTCATCTCAAGCGGCGGCGGCAATCCGAGAGAC 707
 419 GAGCGCGATCTCAAAACCGGCTTCATCTCAAGCGGCGGCGGCAATCCGAGAGAC 476
 708 CAGCGATGACGACTTCATGGAAGAGAGGAGTGGAGAGAGAGAGAGAGAGAGAGAG 766
 477 CAGCGATGACGACTTCATGGAAGAGAGGAGTGGAGAGAGAGAGAGAGAGAGAGAG 536
 767 GAGGGGAGCGGACGAGTTCCTGACGCGGAGCTTCGCG-AGAAGTACGAGGAGTACCA- 824
 537 GAGGGGAGCGG-AGCGAGTTCCTGACGCGGAGCTTCGCGAGAGAGAGAGAGAGAGAG 595
 825 CAGCGAGAGCTTCAGAACATGAGCAA---GCAGAGTTCATCAAGAGAGTACCTGGAAC 880
 596 CAGCGAGAGCTTCAGAACATGAGCAAAGCGGAGAGTTCATCAAGAGAGTACCTGGAAC 655
 881 TGAGAGAGTCTCTGCGCATGAGAGAC--GAGAACACCGGCTGCGGCTGAGAGAGAA 938
 656 CTGAGAGAGTCTCTGCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 715
 939 GCGGCTGGGTGGCGAGAGCGG-CGTGTGCGGAGAGCTGAGCTGAGACCGGCTGC 997

Db 716 GCGGCTGGGTGGCGAGCAGCGGCGGCTGTGCGGAGCTGAGAGCTGGAAGCGGAGCGTTG 775
 QY 998 GCGCGGAGAACCTCCAGCTGCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1040
 Db 776 GCGCGGAGAGCTCCAGTGTGTGAGCAGAGAACCAATGAGCCCG 818

RESULT 11
 BM474968
 LOCUS
 DEFINITION
 AGNCOURT_6476735 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562828
 5', mRNA sequence.
 BM474968
 VERSION
 BM474968.1 GI:18524010
 EST.
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@strs.rii.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12292 row: n column: 13
 High quality sequence stop: 507.
 Location/Qualifiers

FEATURES

source

1. 1294
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5562828"
 /clone_lib="NIH_MGC_88"
 /tissue_type="Duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-Sport6;
 Site: 1; NotI; Site: 2; SalI; Cloned unidirectionally;
 oligo-dt primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."

BASE COUNT

263 a 380 c 407 g 241 t 3 others

Query Match 53.8%; Score 580.6; DB 13; Length 1294;
 Best Local Similarity 89.8%; Pred. No. 2.2e-108;
 Matches 681; Conservative 0; Mismatches 65; Indels 12; Gaps 5;

QY 89 CTGAGCGGCGGCGGAGAGAGCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 148
 Db 1 CTGAGCGGCGGCGGAGAGAGCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
 QY 149 GAGGCTCCCGCAATGTTGGTGGCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 208
 Db 61 GAGGCTCCCGCAATGTTGGTGGCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 QY 209 AACCACTCCCTTCAGACAGCAGGCTGTCCAGAAATCTAGCTGCTGAGAGAGAGAGAG 268
 Db 121 AACCACTCCCTTCAGACAGCAGGCTGTCCAGAAATCTAGCTGCTGAGAGAGAGAGAG 180
 QY 269 AGGCGCAGATGGGAGAGAGCTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 328
 Db 181 AGGCGCAGATGGGAGAGAGAGCTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 QY 329 AACCGAGCGGCGGAGAGAGAGCTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 388
 Db 241 AACCGAGCGGCGGAGAGAGAGCTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

OY		389	TGGGGGCGCTCCGCCAGGGGGCCGAAGAAGATGGGGACACGACAGCAGAGACGTGGGGA	448
Dd		301	TGGGGGCTCTCCGCCAGGGGGCCGAAGAAGATGGGGACACGACAGCAGAGACGTGGGGA	360
OY		449	AGAAAAAACAATAAGAGAGCGCCCGTCCCAAGAAAGACGGCATTTGAAAACCGTACTACAAAC	508
Dd		361	AGAAAAAACAATAAGAGAGCGCCCGTCCCAAGAAAGACGGCATTTGAAAACCGTACTACAAAC	420
OY		509	TGAAC TGGAAGAGAGAGAAAAAGTTTCGACGAGAAAAACAGAGCCTTGAGCTTCAGAGATCC	568
Dd		421	TGACCTGGGAAGAGAGAAAAAGTTTCGACGAGAAAAACAGAGCCTTGAGCTTCAGAGATCC	480
OY		569	GAGCCGACATGTGTCCCCCAAGGGGCCAGCGGTGCCTTATTAACACCACGAG-TTCTCTC	627
Dd		481	GAGCCGACATGTGTCCCCCAAGGGGCCAGCGGTGCCTTATTAACACCACGAGTTTCTCTC	540
OY		628	ATGATGATGATCACGACGACGAGGAGCGGATGTCCTCAAACCGGCTGTACTCCAAGGGGG-	685
Dd		541	ATGATGATGATCACGACGACGAGGAGCGGATGTCCTCAAACCGGCTGTACTCCAAGGGGGC	600
OY		686	CCGCCGCCAAATCC--GACGACACCAGC--GATGACGACTTCATG-----AAGAAGAGG	736
Dd		601	CGCCCGCCAAATCCCGGACGACACACGAGCGGATTGACCACTTCATGGAANCAAGGGGGGT	660
OY		737	CTGAGGAGATGGGGGCGACCGCATGGGATGGGAGGGGACGCGCAGCATTTCTGCACCGGG	796
Dd		661	GGAAGAAAGATGGGGGGGCCAGCGCGATGGGGGAATGGGGAAGGGGGAACGGCGCCCAATT	720
OY		797	ACTTCTCGACGACGATGACGCGGATGCTCCACACGAGAGAGC	834
Dd		721	TTTTCTTCACAGCGGGGGAACTTTCTTCTGTGAAAACC	758
RESULT 12				
AL576112/c				
LOCUS		AL576112	986 bp	mRNA linear EST 16-FEB-2001
DEFINITION		AL576112 LTI_NFL006.PL2 Homo sapiens cdna clone CS0D1072YL06	3	
ACCESSION		prime mRNA sequence.		
VERSION		AL576112		
KEYWORDS		AL576112.1 GI:12937932		
SOURCE		EST.		
ORGANISM		human.		
		Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE		1 (bases 1 to 986)		
AUTHORS		Ll,w.B., Gruber,C., Jesse,J. and Polayes,D.		
TITLE		Full-length cDNA libraries and normalization		
JOURNAL		Unpublished (2001)		
COMMENT		Contact: Genoscope		
		Genoscope - Centre National de Sequencage		
		Bp 191 91006 EVRY cedex - France		
		Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES				
source		Location/Qualifiers		
		1..986		
		/organism="Homo sapiens"		
		/db_xref="taxon:3606"		
		/clone="CS0D1072YL06"		
		/clone_lib="LTI_NFL006.PL2"		
		/issue_type="placenta"		
		/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand CDNA was primed with a NotI-Oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Peng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetechn.com URL : http://fulllength.invitrogen.com"		
BASE COUNT		216 a	284 c	217 g
ORIGIN				265 t
				4 others

Query Match	53.6%	Score 579	DB 9	Length 986
Best Local Similarity	96.7%	Pred. No. 4,5e-108		
Matches 610	Conservative 2	Mismatches 17	Indels 2	Gaps 2
QY	450	GAATAAAGATAGACACCCCTCCACAGAAAGAGCGGATGGAAACCGTACTACAACT	509	
Db	985	GGAAAGAAAAACCTRBAALACGCCGTCCAAAGAGAAAGGCGATT- GAAACCGTACTACAACT	927	
QY	510	GAACTGGGAGAGAGAAAAAAGTTTGAGAGAGAAACAGAGACCTTCAGCTTCAGAGATCCG	569	
Db	926	GACCTGGGAGAGAGAAAAAAGTTTGAGAGAGAAACAGAGACCTTCAGCTTCAGAGATCCG	867	
QY	570	AGCCGAGATGTTCCGCAAGGGCCAGCGGTCGCGCCCTA7AACACACGAGCTTCAT	629	
Db	866	AGCCGAGATGTTCCGCAAGGGCCAGCGGTCGCGCCCTA7AACACACGAGCTTCAT	807	
QY	630	GGATATATACAGACCCAGAGAGCCCGGATTCATAAACCGGCGCTGTACTCCAAAGCGGCGC	689	
Db	806	GGATATATACAGACCCAGAGAGCCCGGATTCATAAACCGGCGCTGTACTCCAAAGCGGCGC	747	
QY	690	CGCCAAATCCGACGACACCCAGATGACGACTTCATGTGAAGAAGGGGTGAGAGATGG	749	
Db	746	CGCCAAATCCGACGACACCCAGATGACGACTTCATGTGAAGAAGGGGTGAGAGATGG	687	
QY	750	GGGACGCATGGGATGGAGGGGACGGCAGCGAGTTTTCAGACGGGCGACTTCGAGAC	809	
Db	686	GGGACGCATGGGATGGAGGGGACGGCAGCGAGTTTTCAGACGGGCGACTTCGAGAC	627	
QY	810	GTACAGAGGGATACCCACGAGAGCGCTTCAGAACATGAGCAAGCGAGGCTCATCAAGA	869	
Db	626	GTACAGAGGGATACCCACGAGAGCGCTTCAGAACATGAGCAAGCGAGGCTCATCAAGA	567	
QY	870	GTACCTGGAAGTGGAGAAATGCTCTCGCGCATGGAGAGACGAGAAACAACCGGCTCGGCT	929	
Db	566	GTACCTGGAAGTGGAGAAATGCTCTCGCGCATGGAGAGACGAGAAACAACCGGCTCGGCT	507	
QY	930	GGAGAGCAAGCGCGGTGGCGAGCAAGCGCGCTGTGGGAGGCTGGAGGCTGGAGCTGGA	989	
Db	506	GGAGAGCAAGCGCGGTGGCGAGCAAGCGCGCTGTGGGAGGCTGGAGGCTGGAGCTGGA	447	
QY	990	CCGGCTCGGCGCCCGAAGCACTCCAGCTCTACCGAGAACGAACTGCACCGGACGACAGA	1049	
Db	446	CC-GGTGGCGCCCGAAGCACTCCAGCTCTACCGAGAACGAACTGCACCGGACGACAGA	388	
QY	1050	GCGAGCGCGCTTTCACAAATTGGAGACTAG	1080	
Db	387	GCGAGCGCGCTTTCACAAATTGGAGACTAG	357	
RESULT 13				
BM474949		1080 bp	mrna	linear
LOCUS	AGNCNCURT_6476734	NIH_MGC_88	Homo sapiens	cdna clone IMAGE:5562756
DEFINITION	5', mRNA sequence.			
ACCESSION	BM474949			
VERSION	BM474949.1	GI:18523991		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	NIH-MGC http://mgc.nci.nih.gov/			
JOURNAL	Unpublished (1999)			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgaabs@remai.nih.gov			
	Tissue Procurement: ATCC			
	CDNA Library Preparation: Life Technologies, Inc.			
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)			
	CDNA Sequencing by: Agencourt Bioscience Corporation			
	CDNA distribution: MGC clone distribution information can be			

|||||
Db 61 AAAATTGACAGAGAAACAGAGCTTCGAGCTTCAAGATCCGACCGAGATGTTGGCA 120
OY 587 AGGCGCAGCGGTCGCGCTTAAACACACGACGATTCCTCATGTGATATCAACGACG 646
Db 121 AGGGCAGCGGTCGCGCTTAAACACACGACGATTCCTCATGTGATATCAACGACG 180
OY 647 AGGACCGGATCTCAAAACCGGCTTACTCCAGCGGCGCGCCAAATCCGACGACA 706
Db 181 AGGACCGGATCTCAAAACCGGCTTACTCCAGCGGCGCGCCAAATCCGACGACA 240
OY 707 CCAGGATGACGACTTCAATGAAGAAGGGGCTGAGAGAGATGGGGCAGCCGATGATG 766
Db 241 CCAGGATGACGACTTCAATGAAGAAGGGGCTGAGAGAGATGGGGCAGCCGATGATG 300
OY 767 GAGGGAGCGGACGAGATTTCTGACGCGGACTTCCGAGACGTACGAGCGGTACCCACA 826
Db 301 GAGGGAGCGGACGAGATTTCTGACGCGGACTTCCGAGACGTACGAGCGGTACCCACA 360
OY 827 CGGAGAGCCTGCGAATATGACCAAGCAGAGCTCATCAAGAGTACCTGGAACGTGAGA 886
Db 361 CGGAGAGCCTGCGAATATGACCAAGCAGAGCTCATCAAGAGTACCTGGAACGTGAGA 420
OY 887 AGTGGCTCTCGGCTATGAGAGCAGAACACCGGCTGCGCTGAGAGCAAGCGGCTGG 946
Db 421 AGTGGCTCTCGGCTATGAGAGCAGAACACCGGCTGCGCTGAGAGCAAGCGGCTGG 480
OY 947 GTGGGAGCAGCGGCTGCGGAGCTGAGAGCTGAGAGCTGGAGCGGCGCGCGAGA 1006
Db 481 GTGGGAGCAGCGGCTGCGGAGCTGAGAGCTGAGAGCTGGAGCGGCGCGCGAGA 540
OY 1007 ACCTCCAGCTGCTGACC 1023
Db 541 ACCTCCAGCTGCTGACC 557

RESULT 15
BE547281 745 bp mRNA linear EST 09-AUG-2000
LOCUS 601073706F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3459960 5',
DEFINITION mRNA sequence.
ACCESSION BE547281
VERSION BE547281
KEYWORDS GI:9775926
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 745)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM8453 row: k column: 01
High quality sequence stop: 730.
Location/Qualifiers
1..745
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3459960"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.4 kb. Library prepared by Life Technologies.
BASE COUNT 195 a 168 c 276 g 106 t
ORIGIN
Query Match 49.9%; Score 539.2; DB 10; Length 745;
Best Local Similarity 93.6%; Pred. No. 5; 6e-100;
Matches 640; Conservative 0; Mismatches 33; Indels 11; Gaps 7;
OY 345 CGAGCTGCTCGCCACGCTTGTGATGACTCCGAGCGCAGTAAAGTTGGGGCTCTGCCGC 404
Db 1 CGAGCTGCTCGCCACGCTTGTGATGACTCCGAGCGCAGTAAAGTTGGGGCTCTGCCGC 58
OY 405 AGGGGCGCAAGAGAGAGTGGGACACACAGAGAGAGTGGGGGAAAGAAACATTAAGAG 464
Db 59 AAGGGGCGCAAGAGAGAGTGGGACACACAGAGAGAGTGGGGGAAAGAAACATTAAGAG 117
OY 465 ACGCCCGTCGAAGAAAGAGCGCATTTGAAACCGTACTACAAAGCTGAACTGGGAAGAA 524
Db 118 ACGCCCGTCGAAGAAAGAGCGCA-TGGAAACCGTACTACAAAGCTGAACTGGGAAGAA 176
OY 525 GAAAGAGTTCCAGAGAAACAGAGCTTTCGAGCTTCAAGAGATCCGAGCCGAGATTTGCC 584
Db 177 GAAAGAGTTCCAGAGAAACAGAGCTTTCGAGCTTCAAGAGATCCGAGCCGAGATTTGCC 236
OY 585 CAAGGCGCAGCGGCTGCGGCTTATACACACAGAGTTCCTCATGATGATCACAGACA 644
Db 237 CAAGGCGCAGCGGCTGCGGCTTATACACACAGAGTTCCTCATGATGATCACAGACA 296
OY 645 GGAGGAGCGGATCTCAAAACCGGCTTACTCCAAAGCGGGCGCGCCCAATCCGACGA 704
Db 297 GGAGGAGCGGATCTCAAAACCGGCTTACTCCAAAGCGGGCGCGCCCAATCCGACGA 355
OY 705 CACGAGCGATGACGACTTTCATGGAAGAAGGGGGTATAGAGATGGGGCAGCGATGGAT 764
Db 356 CACGAGCGATGACGACTTTCATGGAAGAAGGGGGTATAGAGATGGGGCAGCGATGGAT 415
OY 765 GGGAGGGAGCGCAGCGAGTTTCTGCAGCGGAGCTTTCGAGAGCTGACGAGCGGTACCA 824
Db 416 GGGAGGGAGCGCAGCGAGTTTCTGCAGCGGAGCTTTCGAGAGCTGACGAGCGGTACCA 475
OY 825 CACGAGAGCGCTGCGAATGAGCAAGCAGAGCTCATCAAGAGTACCTGGAACCTGA 884
Db 476 CACGAGAGCGCTGCGAATGAGCAAGCAGAGCTCATCAAGAGTACCTGGAACCTGA 535
OY 885 GAATGGCTCTCGCGCATGAGAGAGAGAACACCGGCTGCGGCTGGAGACCAAGCGCT 944
Db 536 GAATGGCTCTCGCGCATGAGAGAGAGAACACCGGCTGCGGCTGGAGACCAAGCGCT 592
OY 945 GGGTGGCGCAGCAGCGGCTGTGGCGGAGCTGAGAGCTGGAGCGCGCTGGCGCGCGCA 1004
Db 593 GGGTGGCGCAGCAGCGGCTGTGGCGGAGCTGAGAGCTGGAGCGCGCTGGCGCGCGCA 649
OY 1005 GAACCTCCAGCTGCTGACCGAGAA 1028
Db 650 GAACCTCCAGCTGCTGACCGAGAA 673

Search completed: June 18, 2003, 01:41:37
Job time : 1832 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 00:13:37 ; Search time 67 Seconds
(without alignments)
4943.446 Million cell updates/sec

Title: US-09-972-758A-1

Perfect score: 1080

Sequence: 1 atggcgcgagccattctgtc.....tttccaagtgtgagactag 1080

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	69	6.4	7218	1	US-08-232-463-14
2	68.4	6.3	1926	4	US-09-249-585A-2
3	68.4	6.3	2580	3	US-09-050-863-2
4	68.4	6.3	2580	4	US-09-359-081-2
5	68.4	6.3	5452	2	US-09-130-114-1
6	68.4	6.3	9600	4	US-08-910-647-1
7	68.4	6.3	9600	4	US-09-620-925-1
8	68.4	6.3	10596	1	US-07-884-811-15
9	68.4	6.3	10596	1	US-07-885-971-15
10	68.4	6.3	10596	1	US-08-087-783A-15
11	68.4	6.3	10596	2	US-08-194-088B-15
12	68.4	6.3	10596	2	US-08-194-087-15
13	68.4	6.3	10596	5	PCT-US93-04648-15
14	67.4	6.2	1926	4	US-09-249-585A-4
15	67.4	6.2	1931	2	US-09-130-114-2
16	65.8	6.1	9551	1	US-08-056-200-93
17	65.8	6.1	9551	2	US-08-800-644-93
18	63.4	5.9	3489	2	US-08-728-323A-1
19	63.4	5.9	3489	2	US-08-298-568-1
20	63.4	5.9	32207	4	US-08-770-379-20
21	63.4	5.9	32207	4	US-08-757-669A-20
22	63.4	5.9	32207	4	US-09-230-371A-20
23	60.2	5.6	2793	1	US-08-209-747-1
24	60.2	5.6	2793	1	US-08-458-298-1
25	58.4	5.4	2313	4	US-09-370-838-157
26	51.2	4.7	397	3	US-09-253-691-3
27	51	4.7	1995	1	US-08-425-069-3

28	51	4.7	1995	2	US-08-317-844B-3	Sequence 3, Appl1
29	50.8	4.7	51259	3	US-08-781-891-209	Sequence 209, App
30	49.6	4.6	3138	4	US-09-434-408-1	Sequence 1, Appl1
31	49	4.5	343	6	5273901-6	Patent No. 5273901
32	49	4.5	16442	3	US-08-781-891-208	Sequence 208, App
33	47.4	4.4	2277	1	US-08-676-967-5	Sequence 5, Appl1
34	47.4	4.4	2277	1	US-08-676-974-5	Sequence 5, Appl1
35	47.4	4.4	2277	2	US-09-098-487-5	Sequence 5, Appl1
36	47.2	4.4	2338	1	US-08-425-069-1	Sequence 1, Appl1
37	47.2	4.4	2338	2	US-08-317-844B-1	Sequence 1, Appl1
38	46.8	4.3	234	1	US-08-469-802B-3	Sequence 3, Appl1
39	46.8	4.3	234	2	US-08-267-803B-3	Sequence 3, Appl1
40	46.6	4.3	1044	1	US-07-975-526-1	Sequence 1, Appl1
41	46.6	4.3	1044	4	US-07-974-409C-422	Sequence 422, App
42	46.6	4.3	1044	4	US-07-974-409C-423	Sequence 423, App
43	46.2	4.3	1618	2	US-08-533-669A-9	Sequence 9, Appl1
44	46.2	4.3	1618	2	US-08-607-509-1	Sequence 1, Appl1
45	46.2	4.3	1618	2	US-08-454-036-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
City: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
CLASSIFICATION: 435
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SRO ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PC-Fls
US-08-232-463-14
Query Match 6.4%: Score 69; DB 1; Length 7218;

Best Local Similarity 3.18; Pred. No. 1.7e-06;
Matches 12; Conservative 288; Mismatches 143; Indels 0; Gaps 0;

```
QY 386 AGTTGGGGCTCTCGCCAGGGGGGAGAGAGAGGGGAGCGACGACGACACTGG 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1453 AGTTAGACGATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1394
QY 446 GGAAGAAAAACATAAGACGCCCTCCAGAAAGAGCGCATTTGAAACCTACTACA 505
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1393 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1334
QY 506 AGTGAAGTGGCAGAGAAAGAAAAAGTTGACGAGAAACAGACCTTCGAGCTTCAAGA 565
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1333 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1274
QY 566 TCCGACCCGAGATGTCGCCAAGGGCCGCGCTCGCCCTTAACACACAGCAGTTC 625
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1273 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1214
QY 626 TCATGATGATCAGCAGCAGAGAGCGCGATCTCAAAACCGCCTGTACTCCAGCGG 685
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1213 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1154
QY 686 CCGCCCAATCCGACGACACGACGAGATGACACTTCATGGAAGAGGGGTGAGAGAG 745
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1153 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1094
QY 746 ATGGGGCAGCGATGGATGGAGGAGGAGCGCA 778
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1093 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1061
```

RESULT 2

```
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: HOTLICK, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2
```

Query Match 6.3%; Score 68.4; DB 4; Length 1926;
Best Local Similarity 44.0%; Pred. No. 1.7e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
QY 406 GGGGGCAGAGAGTGGGGAGCAGACGAGAGAGCTGGGAGAAAAAAGCTAAGAGA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 GGGCAGAGACAGAGAGAGGGGCGACGACGAGAGAGGGGCGACAGAGAGGGCGAG 364
QY 466 CGCCCGTCCAGAAAGAGCGCATTTGAAACCGTACTCAAGCTGAAGTGGGAAGAG 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 GAGGGCAGAGCAGAGAGAGGGGCGACGAGAGAGAGAGGGGCGAGAGAGGGG 424
QY 526 AAAAGTTGACGAGAAAGAGACCTTCAGCTTCAAGATCCGAGCCAGATGTTCCGC 585
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 CAGGACAGAGAGAGGGGCGACGAGAGAGGGGCGACGAGAGGCGAGAGAGAG 484
QY 586 AAGGGCCAGCGCGCTCCGCTATATAACACAGCAGGATTCCTCATGGATGATCAGCGAG 645
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 GGGCAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGGGGCGACGAGAGAGAGAGAG 544
```

QY 646 GAGAGCCGAGATCTCAAAACCGGCTGTACTCCAAAGCGGGCCGCCAAATCCGACGAC 705

Db 545 GAGGGCAGAGCAGAGAGAGAGGGGCGACGAGAGGGGCGACGAGAGAGAGAGAGAGGGG 604

QY 706 ACCAGCGATGACGACTTCATGGAAGAGGGGTGAGAGAGATGGGGCAGCGATGGGATG 765

Db 605 CAGAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGACGAGAGAGGGGCGAGAGAG 664

QY 766 GAGAGGAGCGGACGAGTTTGTGACGGGGACTTGTGAGAGCTGACGAGAGGGGTACCAC 825

Db 665 GAGGGCAGAGGGGCGAGAGGGGCGAGAGCGACGAGAGAGAGAGAGAGAGAGAGAGAG 724

QY 826 ACAGAGCGCTGACGAAATGAGCAAGCAGAGAGCTCATCAGAGAGTACCTGGAAGCTGAG 885

Db 725 CAGAGAGGGCAGAGAGAGAGGGGCGAGAGGGGCGACGAGAGAGAGAGAGAGAGAGAGAG 784

QY 886 AAGTGCCTCTCGCGATGAGAGACGAACCAACCGGCTCGCGCTGAGAGCAAGCGCTG 945

Db 785 GAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 844

QY 946 GGTGGCAGACGCGCGCTGCGGGAGCTGAGAGCTGAGAGCTGAGAGAGCGGCTGCGCGCGAG 1005

Db 845 GAGGGCAGAGAGGGGCGAG 904

QY 1006 AACCTCAGCTGCTGACCGAGAAAGCACTGCAACCGGCGAGAGAGAGAGAGAGCGCG 1059

Db 905 GGGCAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 958

RESULT 3

```
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 6114111
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hsiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/050,863
; FILING DATE: 30-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RWS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-09-050-863-2
```

	Query Match	6.3%	Score 68.4:	DB 3:	Length 2580;
	Similarity	44.0%;	Pred. No. 1.8e-06;		
	Matches	288:	Conservative	0:	Mismatches 366; Indels 0; Gaps 0;
OY	GGGCGCCGAAGGAGGTGGGCACACAGCAGAGCACGCTGGGGAAGAAAAACAATACAGA	465			
Db	GGGCAAGACAGGAGAGAGGGCCAGAGAGAGAGAGGGGCCAGAGAGGGCCAG	747			
OY	CGCCCCGTCCAAAGAAAGCGGCAATTGGMAAACCCTACTACAAAGTGAACCTGGGAAGAA	525			
Db	GAGGGCGCAGAGCAGAGAGAGGGCCAGAGCAGAGAGAGGGGCCAGAGAGGGCCAGAGAGGGG	807			
OY	AAAAGTTTCACGAGAAACAGAGCCCTTGACCTTCMAAGATCCAGGCCGAGATGTCGCC	585			
Db	CAGAGACAGAGGAGGGGCAGAGACAGAGAGAGGGGGCAGAGGGGCCAGAGACAGAGAGAG	867			
OY	AAGGGCCAGCGCGTGGCGCCCTATACACACAGCATGTTCTCATAGATGATCAGACACAG	645			
Db	GGGCAGAGAGGGCCAGAGAGGGCCAGAGCAGAGAGAGGGGCCAGAGACAGAGAGAGGGCAG	927			
OY	GAGGAGCCGGATCTCAAACCCGGCCTGTACTCCAAGCGGGCCGCCCAATTCGACGAC	705			
Db	GAGGGCCAGAGCAGAGAGAGGGCCAGAGAGGGCCAGAGAGGGGCCAGAGACAGAGAGAGGGG	987			
OY	ACCAGCGATACGACATTCATGAAAGAGGGGTTAGACGATGGGGCCAGCGATGGATG	765			
Db	CAGGACGACGAGGGGGCAGAGAGGGGCAGAGAGGGCCAGAGCAGAGACAGACAGACAGAG	1047			
OY	GGAGGGGACGGCACACGASTTTCTCAGCGGGACCTTCTCGAGACGTAACGACGGTACCAC	825			
Db	GAGGGGCAGGGGGCCAGCAGGGCCAGAGAGGAGGAGGGGCCAGAGAGAGGGCCAGAGAG	1107			
OY	ACGAGAGACCTTCAGAACATGAGCAAGCAGAGCTCATCAAGAGTAACCTGGAACACTGGAG	885			
Db	CAGGAGGGGGCAGAGAGCAGAGAGGGCCAGAGAGGGGCCAGAGCAGAGAGGGGCCAGAGAGGGCCAG	1167			
OY	AAGTGCCTCTTCGCCCATATGAGAGAGCAACAACCGCTCGCGTGTGAGAGCAACCGCGCTG	945			
Db	GAGCAGAGAGGGGCAGAGAGGGCCAGAGAGGGCCAGAGAGGGGCCAGAGAGGGGCCAGAGAG	1227			
OY	GGTGGCCACGACGGCGCTGTCTGGAGCTGAGAGCTGGAAGCTGGAACCGGCTGCGGGCCAG	1005			
Db	GAGGGGCAGGGGGCCAGAGAGGGCCAGAGAGGGGCCAGAGAGGGGCCAGAGAGGGGCCAGAGAG	1287			
OY	AACCTCCAGCTGTGACCCGAGAACGAACACTGACCGGCGAGCGAGCGAGCCGCG	1059			
Db	GGGCAAGAGCAGAGAGGGCCAGAGAGGGGCCAGAGAGGGGCCAGAGAGGGGCCAGAGAGGGCCAG	1341			
RESULT 4					
US-09-359-081-2					
Sequence 2, Application US/09359081					
Patent No. 6316223					
GENERAL INFORMATION:					
APPLICANT: Lao, Ying					
Huang, Betty					
Payan, Don					
TITLE OF INVENTION: Mammalian Protein Interaction Cloning					
NUMBER OF SEQUENCES: 5					
CORRESPONDENCE ADDRESS:					
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert					
STREET: 4 Embarcadero Center, Suite 3400					
CITY: San Francisco					
STATE: CA					
COUNTRY: USA					
ZIP: 94111-4187					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Floppy disk					
COMPUTER: IBM PC compatible					
OPERATING SYSTEM: PC-DOS/MS-DOS					
SOFTWARE: PatentIn Release #1.0, Version #1.30					
CURRENT APPLICATION DATA:					

```

1  APPLICATION NUMBER: US/09/359,081
2  FILING DATE: 22-Jul-1999
3  CLASSIFICATION: <Unknown>
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER: 09/050,863
6  FILING DATE: <Unknown>
7  ATTORNEY/AGENT INFORMATION:
8  NAME: Silva, Robin M.
9  REGISTRATION NUMBER: 38,304
10 REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: (415) 781-1989
13 TELEFAX: (415) 949-8711
14
15 INFORMATION FOR SEQ ID NO: 2:
16 SEQUENCE CHARACTERISTICS:
17     LENGTH: 2580 base pairs
18     TYPE: nucleic acid
19     STRANDEDNESS: unknown
20     TOPOLOGY: unknown
21
22 MOLECULE TYPE: DNA
23
24 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
25 US-09-359-081-2
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
9
```

US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Hotlick, Robert A.
; APPLICANT: Dama, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Epilomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130.114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
; US-09-130-114-1

Query Match 6.3%; Score 68.4; DB 2; Length 5452;
Best Local Similarity 44.0%; Pred. No. 2, 1e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGCGAAGAGAGTGGGACAGCAGACAGCTGGGGAAGAAAACTAAGAGA 465
DB 2117 GGGCAGAGAGCAGAGAGGGGCGAGCAGAGAGAGGAGGCGAGAGGGGCGAG 2058
QY 466 CGCCCGTCCAAAGAAAGCGGCGCTTGAACCGTACTACAGCTGAACCTGGGAAG 525
DB 2057 GAGGGCAGAGAGAGAGAGGGGCGAGCAGAGAGAGAGAGGGGCGAGAGGGG 1998
QY 526 AAAAGTTGACAGCAAGAACAGACCTTCGAGCTTCAAGATCCAGCCGAGATGTTCC 585
DB 1997 CAGGACAGAGAGAGGGGCGAGCAGAGAGAGAGGCGAGAGGGGCGAGAGAG 1938
QY 586 AAGGGCAGCGCGCTCCGCGCTTAACACACAGCAGTCTCTCATGTGATGACAGAC 645
DB 1937 GGGCAGAGAGAGAGAGGGGCGAGCAGAGAGAGAGGGGCGAGAGAGAGGGGCGAG 1878
QY 646 GAGGACCGGATCTCAAAACCGGCTGTACTCTCAAGAGGGGCGCCCGCAATCCGAG 705
DB 1877 GAGGGCAGAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGAGAGGGG 1818
QY 706 ACCAGCGATGACACTTTCATGAAGAAGGGGTGAGAGAGATGGGGCGACGATGGGATG 765
DB 1817 CAGGACAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAG 1758
QY 766 GAGGGGAGCGGCGAGCGAGTTTTCGACGCGGAGCTTTCGAGACGTRACAGCGGTAC 825
DB 1757 GAGGGGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGAGAGAG 1698
QY 826 ACGGAGCGCTTGAGAAATGAGCAAGCAGAGCTCATCAAGAGTAACTGGAATGGAG 885
DB 1697 CAGGAGGGCAGAGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
QY 886 AAGTGCCTTCGCGCATGAGAGAGCAACACCGGCTGCGGCTGAGAGCAACCGGCTG 945
DB 1637 GAGCAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
QY 946 GGTGGCAGAGAGCGCGCTGTGCGGAGCTGTGAGAGCTGAGAGCGGCTGCGCGAG 1005
DB 1577 GAGGGGAGAGAGGGGCGAGAGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAG 1518
QY 1006 AACCTCAGCTGTGAGACCGAGAACTGACCGGAGCAGAGAGAGAGAGAGAGAGAG 1059
DB 1517 GGGCAGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464

RESULT 6
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433

GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-910-647-1

Query Match 6.3%; Score 68.4; DB 4; Length 9600;
Best Local Similarity 44.0%; Pred. No. 2, 4e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGCGAAGAGAGTGGGACAGCAGACAGCAGCTGGGGAAGAAAACTAAGAGA 465
DB 734 GGGCAGAGAGCAGAGAGGGGCGAGCAGAGAGAGAGGGGCGAGAGAGGGGCGAG 793
QY 466 CGCCCGTCCAAAGAAAGCGGCGCTTGAACCGTACTACAGCTGAACCTGGGAAG 525
DB 794 GAGGGCAGAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGAGAGGGG 853
QY 526 AAAAGTTGACAGCAAGAACAGACCTTCGAGCTTCAAGATCCAGCCGAGATGTTCC 585
DB 854 CAGGACAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGGGGCGAGAGAGAGAG 913
QY 586 AAGGGCAGCGCGCTCCGCGCTTAACACACAGCAGTCTCTCATGTGATGACAGAC 645
DB 914 GGGCAGAGAGGGGCGAGAGGGGCGAGCAGAGAGAGAGGGGCGAGAGAGAGGGGCGAG 973
QY 646 GAGGACCGGATCTCAAAACCGGCTGTACTCTCAAGAGGGGCGCCCGCAATCCGAG 705
DB 974 GAGGGCAGAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGAGAGAGAGAG 1033
QY 706 ACCAGCGATGACACTTTCATGAAGAAGGGGTGAGAGAGATGGGGCGACGATGGGATG 765
DB 1034 CAGGACAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAG 1093
QY 766 GAGGGGAGCGGCGAGCGAGTTTTCGACGCGGAGATTCGGAAGCGTRACAGCGGTAC 825
DB 1094 GAGGGGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1153
QY 826 ACGGAGCGCTTGAGAAATGAGCAAGCAGAGAGCTCATCAAGAGTAACTGGAATGGAG 885
DB 1154 CAGGAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAG 1213

QY	886	AAGGCTCTTCGGCCATGGAGGAGGAGCAAAACGGCGTGGGGTGGAGAGCAAGCGGCTG	945
Db	1214	GAGCAGGAGGGGGCAGGAGGGGCAAGGACCAGGAGGAGGGCCAGGAGGGCCAGAGCTAGGAG	1273
QY	946	GGTGGCCACGACCGCGCGCTGTCCGGGAGCGTGAGCTGGAGCTCGGCGCTCGCGCCAG	1005
Db	1274	GAGGGGCAAGGAGGGGCAAGAACAGGAGGGGCAGGAGGGGCAGGAGCAAGAGGGGCAAGGAG	1333
QY	1006	AACCTCCAGCTGCTGACCCGAGAACGAACTGCACCGGCGAGCGAGGACGAGCGCCG	1059
Db	1334	GGGCAAGGACGAGGAGGGGCAAGGAGGGCCAGGAGCAAGGAGGAGGGGCAAGGACAG	1387

RESULT 7
UC-09-63

```

US-09-620-925-1
: Sequence 1, Application US/09620925
: Patent No. 6468986
: GENERAL INFORMATION:
: APPLICANT: Zuckermann et al.
: TITLE OF INVENTION: Compositions and Methods for
: Polynucleotide Delivery
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/620,925
: FILING DATE: 21-Jul-2000
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/910,647
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Fujita, Sharon M.
: REGISTRATION NUMBER: 38,459
: REFERENCE/DOCKET NUMBER: 1218.002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 923-2706
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9600 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 1:
: TS-09-620-925-1

```

[illegible]

QY	586	AAGGGCCACCGCGGTGCGCCCTATTAAACACACGCGCTTTCATGATGATGCACACCG	645
Db	914	GGCGACGAGGGGGCCAGAGAGGGCCAGACGACGAGAGGGGCCAGAGACAGAGGAGGGCCAG	973
QY	646	GAGGAGCCCGGATCTCCAAAACCGCCCTGTACTCCAAAGCGGGCCGCCCAATCCGACGAC	705
Db	974	GAGGGCGAGGACGAGAGAGAGGGGCGAGGAGGGCGAGAGGGCCAGACAGAGAGAGGGG	1033
QY	706	ACCAAGCCATTACGACTTCATGGAAGAAAGGGGTGAGAGATGGGGGCCAGCATGGGATG	765
Db	1034	CAGGAGCAGGAGGGGGCAGAGGCGGCAGAGAGGGGCCGAGACGAGAGAGGGGCGAGGACGAG	1093
QY	766	GGAGGGGACGCGACGAGATTTCTCGACGGGGACTTCTGAGACCTACGAGCGGTACACAC	825
Db	1094	GAGGGGCCAGAGGGGGCAGAGGGGCGAGAGCCAGAGAGGGGCCAGAGACAGAGAGGGGCCAGAG	1153
QY	826	ACGAGAGACCTTCGAGAACATGACCAAGCAGAGACTCATCAAGGATACCTGGAACTCGAG	885
Db	1154	CAGGAGGGGGCCAGGAGCAGAGAGGGGCGAGAGGGGCCAGGACAGAGCGGCGAGAGAGGGGGCAG	1213
QY	886	AAGTGTCTCTGCGCATGGAGGACGAGAAACCGGCTGCGGCTGGAGAGCAAGGGGGCTG	945
Db	1214	GAGCAGGAGGGGGCCAGAGAGGGGCCAGGACGAGGAGGAGGGGCCAGAGAGGCGACGACGAG	1273
QY	946	GGTGGCGACGACGCGCGTGTCCGGAGCTGTGAGCTTGACACGGCTGCGCCCGAG	1001
Db	1274	GAGGGGACGAGAGGGGCGACGAGACGAGGAGGGGCGAGAGGGGCCAGAGCAGAGAGGGCGAGAG	1333
QY	1006	AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCGACGAGAGCAAGCCCG	1059
Db	1334	GGGAGGAGCGAGAGGGGCGAGAGCGGACGAGGAGGAGAGAGGGCGAGAGGAG	1387

RESULT 8

```

US-07-884-811-15
: Sequence 15, Application US/07884811
: Patent No. 5316921
:
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R
: TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patln (genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/884,811
: FILING DATE: 19920518
: CLASSIFICATION: 350
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dreger, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: 755.1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-3216
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ. ID NO.: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: NUCLEIC ACID

```


